523
9.11.76

Journal of Personality and Social Psychology

Editor: Daniel Katz, *University of Michigan*
Associate Editor: Robert B. Zajonc, *University of Michigan*

Consulting Editors

Robert P. Abelson
Yale University
Joseph B. Adelson
University of Michigan
Gordon W. Allport
Harvard University
Norman H. Anderson
University of California, San Diego
Elliot Aronson
University of Texas
Albert Bandura
Stanford University
Leonard Berkowitz
Cornell University
James Bieri
University of Texas
Jack W. Brehm
Duke University
Donald R. Brown
University of Michigan
Eugene Burnstein
University of Michigan
Arnold H. Buss
Rutgers University
Launor F. Carter
Systems Development Corporation
Rosalind D. Cartwright
University of Illinois Medical Center, Chicago
Dwight W. Chapman, Jr.
Vassar College
Irvin L. Child
Yale University
Nickolas B. Cottrell
Kent State University
William Dember
University of Cincinnati
Morton Deutsch
Columbia University
Seymour Feshbach
University of Colorado
Donald W. Fiske
University of Chicago
Basil S. Georgopoulos
University of Michigan
Edwin P. Hollander
State University of New York at Buffalo
Murray Horwitz
New York University
Herbert C. Kelman
University of Michigan
Daniel J. Levinson
Connecticut Mental Health Center New Haven
Richard A. Littman
University of Oregon
Robert B. MacLeod
Cornell University
Melvin Manis
Veterans Administration Hospital, Ann Arbor, Michigan
David C. McClelland
Harvard University
Charles G. McClintock
University of California, Santa Barbara
William J. McGuire
Columbia University
Vincent Nowlis
University of Rochester
Harold L. Raush
University of Michigan
Milton E. Rosenbaum
University of Iowa
William A. Scott
University of Colorado
Janet Taylor Spence
University of Texas
Percy H. Tannenbaum
University of Wisconsin
Harry C. Triandis
University of Illinois
Joseph Veroff
University of Michigan

Helen Orr
Managing Editor

Joanne Bergen
Technical Editor

Henry S. Salmon
Advertising and Business Manager

Robert D. Gorman
Subscription Manager

VOLUME 6
1967

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Copyright © 1967 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 6

ARTICLES, BRIEF ARTICLES

Abrahams, Darcy. See Walster, Elaine.
Andrews, John D. W. The Achievement Motive and Advancement in Two Types of Organizations 163
Aronson, Vera. See Walster, Elaine.
Atkins, Alvin L., Deaux, Kay K., & Bieri, James. Latitude of Acceptance and Attitude Change: Empirical Evidence for a Reformulation 47
Balloun, Joe L. See Brock, Timothy C.
Bass, Alan R. See Komorita, S. S.
Bauer, Raymond A. See Gergen, Kenneth J.
Baumal, Evelyn. See Watson, David.
Bernard, J. L., & Eisenman, Russell. Verbal Conditioning in Sociopaths with Social and Monetary Reinforcement 203
Berscheid, Ellen, & Walster, Elaine. When Does A Harm-Doer Compensate a Victim? 435
Berscheid, Ellen. See Walster, Elaine.
Bieri, James. See Atkins, Alvin L.
Bishop, Frances V. The Anal Character: A Rebel in the Dissonance Family 23
Bobrove, Philip H. See Cowen, Emory L.
Bootzin, Richard R., & Stephens, Mark W. Individual Differences and Perceptual Defense in the Absence of Response Bias 408
Brehm, Jack W. See Jones, Russell A.
Brock, Timothy C., & Balloun, Joe L. Behavioral Receptivity to Dissonant Information 413
Bryan, James H., & Test, Mary Ann. Models and Helping: Naturalistic Studies in Aiding Behavior 400
Buss, Arnold H., & Portnoy, Norman W. Pain Tolerance and Group Identification 106
Butzine, Kent W. See Insko, Chester A.
Collins, Barry E. See Helmreich, Robert L.
Coombs, Clyde H. Thurstone's Measurement of Social Values Revisited Forty Years Later 85
Cooper, Joel. See Linder, Darwyn E.
Cottrell, Nickolas B., & Wack, Dennis L. Energizing Effects of Cognitive Dissonance Upon Dominant and Subordinate Responses 132
Cowen, Emory L., Bobrove, Philip H., Rockway, Alan M., & Stevenson, John. Development and Evaluation of an Attitudes to Deafness Scale 183
Davis, Henry P. See Dustin, David S.
Deaux, Kay K. See Atkins, Alvin L.
Dittmann, Allen T., & Llewellyn, Lynn G. The Phonemic Clause as a Unit of Speech Decoding 341
Ditrichs, Raymond, Simon, Seymore, & Greene, Barry. Effect of Vicarious Scheduling on the Verbal Conditioning of Hostility in Children 71
Druckman, Daniel. Dogmatism, Prenegotiation Experience, and Simulated Group Representation as Determinants of Dyadic Behavior in a Bargaining Situation 279
Dustin, David S., & Davis, Henry P. Authoritarianism and Sanctioning Behavior 222
Dworkin, Earl S., & Efran, Jay S. The Angered: Their Susceptibility to Varieties of Humor 232
Earle, Margaret J. Bilingual Semantic Merging and an Aspect of Acculturation 304
Ebbe, Christopher. See Peterson, David E.
Efran, Jay S. See Dworkin, Earl S.
Eisenman, Russell. See Bernard, J. L.
Feather, N. T. Level of Aspiration and Performance Variability 37
Feshbach, Norma D. Nonconformity to Experimentally Induced Group Norms of High- versus Low-Status Members 55
Fischer, Donald, & Rule, Brendan G. Anti-Semitism, Stress, and Anchor Effects on Interpersonal Judgments 447
Ford, LeRoy H., Jr., & Hersen, Michel. Need Approval, Defensive Denial, and Direction of Aggression in a Failure-Frustration Situation 229
Foster, Robert. See Ryan, E. Dean.
Frankie, Gary. See Hetherington, E. Mavis.
Geer, James H., & Turteltaub, Alan. Fear Reduction Following Observation of a Model 327
Gerard, Harold B. Compliance, Expectation of Reward, and Opinion Change 360
Gergen, Kenneth J., & Bauer, Raymond A. Interactive Effects of Self-Esteem and Task Difficulty on Social Conformity 16
Georgas, James G. See Vassiliou, Vasso.
Goldrich, Judith March. A Study in Time Orientation: The Relation Between Memory for Past Experience and Orientation to the Future 216
Gould, Laurence. See Zigler, Edward.

Greene, Barry. See Ditrichs, Raymond.
Griswold, Barbara B. See Mendelsohn, Gerald A.
Hartup, Willard W. See Shultz, Thomas R.
Helmreich, Robert L., & Collins, Barry E. **Situational Determinants of Affiliative Preference under Stress** 79
Hersen, Michel. See Ford, Leroy H., Jr.
Hetherington, E. Mavis, & Frankie, Gary. **Effects of Parental Dominance, Warmth, and Conflict on Imitation in Children** 119
Hoffman, L. Richard, & Maier, Norman R. F. **Valence in the Adoption of Solutions by Problem-Solving Groups: II. Quality and Acceptance as Goals of Leaders and Members** 175
Holzman, Philip S. See Rousey, Clyde.
Hunt, Raymond G., & Lin, Tip Kan. **Accuracy of Judgment of Personal Attributes from Speech** 451
Insko, Chester A., & Butzine, Kent W. **Rapport, Awareness, and Verbal Reinforcement of Attitude** 225
Jellison, Jerald M. See Mills, Judson.
Jones, Edward E. See Linder, Darwyn E.
Jones, Russell A., & Brehm, Jack W. **Attitudinal Effects of Communicator Attractiveness when One Chooses to Listen** 64
Jones, Stephen C., & Ratner, Carl. **Commitment to Self-Appraisal and Interpersonal Evaluations** 442
Karlins, Marvin. **Conceptual Complexity and Remote-Associative Proficiency as Creativity Variables in a Complex Problem-Solving Task** 264
Katkin, Edward S. See Waterman, Caroline K.
Komorita, S. S., & Bass, Alan R. **Attitude Differentiation and Evaluative Scales of the Semantic Differential** 241
Komorita, S. S., & Mechling, John. **Betrayal and Reconciliation in a Two-Person Game** .. 349
Lazarus, Richard S. See Opton, Edward M., Jr.
Lebovits, Binyamin Z., & Ostfeld, Adrian M. **Personality, Defensiveness, and Educational Achievement** 381
Leventhal, Howard, Watts, Jean C., & Pagano, Francia. **Effects of Fear and Instructions on How to Cope with Danger** 313
Levine, Jacob. See Zigler, Edward.
Levy, Leon H. **Awareness, Learning, and the Beneficient Subject as Expert Witness** 365
Lin, Tip Kan. See Hunt, Raymond G.
Linder, Darwyn E., Cooper, Joel, & Jones, Edward E. **Decision Freedom as a Determinant of the Role of Incentive Magnitude in Attitude Change** 245
Lipsitt, Paul D., & Strodtbeck, Fred L. **Defensiveness in Decision Making as a Function of Sex-Role Identification** 10
Llewellyn, Lynn G. See Dittmann, Allen T.
Lowin, Aaron. **Approach and Avoidance: Alternate Modes of Selective Exposure to Information** 1
Maier, Norman R. F. See Hoffman, L. Richard.
Marks, Edmond. **Some Situational Correlates of Recognition-Response Level** 102
McBride, Loren, & Moran, Gary. **Double Agreement as a Function of Item Ambiguity and Susceptibility to Demand** 115
McClelland, Lucille. See Richey, Marjorie H.
Mechling, John. See Komorita, S. S.
Mehrabian, Albert, & Wiener, Morton. **Decoding of Inconsistent Communications** 109
Mendelsohn, Gerald A., & Griswold, Barbara B. **Anxiety and Repression as Predictors of the Use of Incidental Cues in Problem Solving** 353
Miller, Ralph R. **No Play: A Means of Conflict Resolution** 150
Mills, Judson, & Jellison, Jerald M. **Effect on Opinion Change of How Desirable the Communication is to the Audience the Communicator Addressed** 98
Moran, Gary. See McBride, Loren.
Morgan, William R., & Sawyer, Jack. **Bargaining, Expectations, and the Preference for Equality over Equity** 139
Murdoch, Peter. **Development of Contractual Norms in a Dyad** 206
Newcomb, Thodore, Jr. See Zander, Alvin.
Norman, Martin. See Schaefer, Judith B.
Oakes, William F. **Verbal Operant Conditioning, Intertrial Activity, Awareness, and the Extended Interview** 198
Oakes, William F. See Thaver, Falak.
Opton, Edward M., Jr., & Lazarus, Richard S. **Personality Determinants of Psycholphysiological Response to Stress: A Theoretical Analysis and an Experiment** 291
Ostfeld, Adrian M. See Lebovits, Binyamin Z.
Pagano, Francia. See Leventhal, Howard.
Peterson, David E., Saltzstein, Herbert D., & Ebbe, Christopher. **Sequential Effects in Social Influence** 169
Portnoy, Norman W. See Buss, Arnold H.

RATNER, CARL. See JONES, STEPHEN C.
REIMANIS, GUNARS. Increase in Psychological Anomie as a Result of Radical and Undesirable Change Expectancy 454
RICHEY, MARJORIE H., MCCLELLAND, LUCILLE, & SHIMKUNAS, ALGIMANTAS M. Relative Influence of Positive and Negative Information in Impression Formation and Persistence 322
ROCKWAY, ALAN M. See COWEN, EMORY L.
ROUSEY, CLYDE, & HOLZMAN, PHILIP S. Recognition of One's Own Voice 464
RULE, BRENDAN G. See FISCHER, DONALD.
RYAN, E. DEAN, & FOSTER, ROBERT. Athletic Participation and Perceptual Augmentation and Reduction 472
SALTZSTEIN, HERBERT D. See PETERSON, DAVID E.
SAWYER, JACK. See MORGAN, WILLIAM R.
SCHAEFER, JUDITH B., & NORMAN, MARTIN. Punishment and Aggression in Fantasy Responses of Boys with Antisocial Character Traits 237
SHIMKUNAS, ALGIMANTAS M. See RICHEY, MARJORIE H.
SHULTZ, THOMAS R., & HARTUP, WILLARD W. Performance under Social Reinforcement as a Function of Masculinity-Femininity of Experimenter and Subject 337
SIMON, SEYMORE. See DITRICHS, RAYMOND.
SMITH, SEWARD. See WHEELER, LADD.
STEPHENS, MARK W. See BOOTZIN, RICHARD R.
STEVENSON, JOHN. See COWEN, EMORY L.
STEWART, RALPH H. Birth Order and Dependency 192
STRODTBECK, FRED L. See LIPSITT, PAUL D.
TEST, MARY ANN. See BRYAN, JAMES H.
THAYER, FALAK, & OAKES, WILLIAM F. Generalization and Awareness in Verbal Operant Conditioning 391
TURTELTAUB, ALAN. See GEER, JAMES H.
VALINS, STUART. Emotionality and Information Concerning Internal Reactions 458
VASSILIOU, GEORGE. See VASSILIOU, VASSO.
VASSILIOU, VASSO, GEORGAS, JAMES G., & VASSILIOU, GEORGE. Variations in Manifest Anxiety Due to Sex, Age, and Education 194
WACK, DENNIS L. See COTTRELL, NICKOLAS B.
WALSTER, ELAINE, BERSCHEID, ELLEN, ABRAHAMS, DARCY, & ARONSON, VERA. Effectiveness of Debriefing Following Deception Experiments 371
WALSTER, ELAINE. See BERSCHEID, ELLEN.
WATERMAN, CAROLINE K., & KATKIN, EDWARD S. Energizing (Dynamogenic) Effect of Cognitive Dissonance on Task Performance 126
WATSON, DAVID. The Relationship between Locus of Control and Anxiety 91
WATSON, DAVID, & BAUMAL, EVELYN. Effects of Locus of Control and Expectation of Future Control Upon Present Performance 212
WATTS, JEAN C. See LEVENTHAL, HOWARD.
WEINSTEIN, LAURA. Social Experience and Social Schemata 429
WHEELER, LADD, & SMITH, SEWARD. Censure of the Model in the Contagion of Aggression .. 93
WIENER, MORTON. See MEHRABIAN, ALBERT.
WYER, ROBERT S., JR. Behavioral Correlates of Academic Achievement: Conformity under Achievement- and Affiliation-Incentive Conditions 255
ZANDER, ALVIN, & NEWCOMB, THEODORE, JR. Group Levels of Aspiration in United Fund Campaigns 157
ZIGLER, EDWARD, LEVINE, JACOB, & GOULD, LAURENCE. Cognititve Challenge as a Factor in Children's Humor Appreciation 332

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 1-9

APPROACH AND AVOIDANCE:

ALTERNATE MODES OF SELECTIVE EXPOSURE TO INFORMATION[1]

AARON LOWIN

Iowa State University

A model of selective exposure was developed and tested. The model considered both the possibilities of approach and of avoidance of either consonant or dissonant information, especially as a function of message (ease of refutation) and actor (confidence) variables. It predicted that for messages hard to refute consonant material would be preferred to dissonant; the reverse should obtain for messages easy to refute. Data were presented from studies conducted during the 1964 presidential campaign. Nearly 900 undergraduates were mailed camouflaged but realistic offers of political propaganda. The main study verified the predictions, but a 2nd study supported only some of the predictions.

A recent flurry of reports has laid to rest the popular belief that people generally tend to avoid messages with which they disagree. In their 1965 selective-exposure-literature review Freedman and Sears wrote ". . . experimental evidence does not demonstrate that there is a general psychological tendency to avoid nonsupportive . . . information [p. 69]." The inconsistent pattern of findings (cf. Freedman & Sears, 1965; Lowin, 1965) suggests the presence of mediating variables which shape selective-exposure processes. This paper focuses on two such variables: *actor confidence* and *ease of message refutation.*

THEORY

Why, one wonders, should a person necessarily try to avoid a dissonant message? An individual who has a stake in maintaining a belief structure may choose to deliberately approach dissonant information in order to try to refute that information. The successful refutation of a dissonant cognition is probably as functional for belief-system stability as is its avoidance. Challenging dissonant elements is, however, not always an attractive strategy, for the person is limited in the amount of evidence he can marshal in support of his own views. If a discordant belief is approached and if in the ensuing conflict that cognition proves more potent than was initially presumed, the actor will find himself experiencing the very dissonance he would have preferred to alleviate, or at least ignore. By indiscriminately approaching all dissonant information, the actor may often emerge holding a position more difficult than that with which he started.

The choice between approaching and avoiding dissonant information should by this analysis depend on the confidence the actor feels vis-à-vis the message. Two variables—confidence in own opinion (an actor parameter) and estimated ease of message refutation (a message parameter)—may well, then, shape response to dissonant information. Approach is most likely when the actor is confident of his position and when he considers the dissonant message to be a weak one. Avoidance most probably occurs when an unsure actor identifies a strongly dissonant message. The analyses reported below deal primarily with the message variable. It is predicted that, whereas strong dissonant messages will be avoided, weak ones will be approached.

Consider now that subset of belief systems in which social consensus is the ultimate source of validity. Two such instances come to mind. In the absence of objective criteria, people often rely on the opinions of others as yardsticks for their own beliefs (for instance, in matters of aesthetics or competence; see

---

[1] This paper is a partial summary of a dissertation submitted to the Department of Social Psychology, Columbia University. For his invaluable support I am grateful to J. Stacy Adams. I am indebted to many persons for their generous assistance in locating or interviewing subjects or in analyzing data. Space considerations alone preclude their mention here. This research, carried out while the author was a National Science Foundation Fellow, was supported by a grant from Columbia University.

Festinger, 1954). There are also phenomena whose final disposition is inherently linked with popularity (fads, social protests, political beliefs—especially elections). For present purposes, what is important about this class of belief systems is that each actor draws support for his own system from the fact that other persons agree with his stand. Imagine now that the actor becomes aware that someone else who agrees with him does so for very flimsy, easily refuted reasons. Although ostensibly consonant with the actor's views, these reasons may be uncomfortable in that, under attack, persons accepting them may change their minds, thus withdrawing social support from the actor's belief system. It is suggested here that in a controversial realm in which social support plays a central role weak consonant information may prove uncomfortable. The mediating effect of ease of message refutation on selective exposure can now be extended to consonant messages: (under the above constraint) whereas strong consonant information is approached, weak consonant information is avoided.

If preference data for strong and weak consonant and dissonant information are compared, the above analyses suggest the following hypotheses. Among messages of equal general interest:

1. Strong consonant messages are preferred to weak consonant ones;
2. weak dissonant messages are preferred to strong dissonant ones;
3. strong consonant messages are preferred to strong dissonant ones;
4. weak dissonant messages are preferred to weak consonant ones.

The fourth hypothesis is an especially strategic one, for it alone discriminates between the present approach and the traditional one. The traditional argument states that consonant information is always approached and dissonant always avoided. The more potent the message, the stronger the approach or avoidance tendency; the weaker the message, the weaker the tendency. The present position, on the other hand, suggests an interaction of congruence with message strength. Strong consonant and weak dissonant messages are preferred to weak consonant or strong dissonant. The first three hypotheses above may be derived from either position. But the fourth hypothesis discriminates between the two. If weak consonant messages are merely weakly approached and weak dissonant ones weakly avoided, then subjects should always prefer weak consonant to weak dissonant. At worst, the two should prove of equal interest. But, if weak consonant messages are indeed avoided and weak dissonant ones approached, then subjects should clearly prefer weak dissonant to weak consonant arguments. The examination of this prediction was the central intent of the following two studies.

METHOD

During the Johnson-Goldwater presidential campaign of 1964, realistic propaganda of known position and ease of refutation was offered to persons whose political views had earlier been ascertained. Carefully designed samples of brochures were mailed to experimentally unsensitized subjects who could order the free material by returning reply-paid postcards.

The brochures being offered were represented by a series of arguments. The arguments were either easy or hard to refute and either consonant or dissonant with the subject's own views. The pertinent characteristics of the arguments were known, as they were chosen for just these characteristics from a much larger item pool evaluated by judges.

The studies were set in a political campaign because political selective-exposure studies have met with more success than have others, and because the stability of political attitudes permitted us to discount any possible persuasive effects of the message on the attitudes of the subjects. The experiment was staged so that, to a subject, it should appear that his name had somehow been included on the mailing list of one of the myriad propaganda groups which spring to life during the course of a major election campaign. Subjects were thus unaware of any experimental intent in the letters they received. That this was, in fact, universally the case was ascertained in postexperimental interviews. This aspect of the study was deliberately introduced so that intellectualization (a possible leveling factor in free-choice selectivity research) would be minimized. Who, we asked, sends for what?

EXPERIMENT I

Subjects in this experiment received one of four letters. The letters were identical in all respects except the critical stimuli (samples of the brochure being offered). The critical stimuli followed two introductory paragraphs which identified the organization, its stand, and its purpose in making campaign literature available. Each source organization was partisan and made no secret of the fact that it was interested in "informing the reader of the facts" so that he would vote for their candidate on election day. The brochure samples then followed. Two final paragraphs summarized the offer and urged its acceptance. A recipient of such an offer was not given a choice among various messages. He could respond to a specific offer or choose to ignore that offer.

The only difference between messages lay in the stimuli. These were, of course, deliberately selected to favor either one candidate or the other, and to be either maximally easy or hard to refute. Each stimulus consisted of six items, three worded positively and three negatively, arranged in alternate order.

Sources of Items

Items were collected from many sources supporting or attacking the various candidates. These ranged from Communist party publications to those of the Patrick Henry League, a right-wing organization. A final pool of 48 pro-Johnson, 48 con-Johnson, 48 pro-Goldwater, and 48 con-Goldwater items, all different in content, resulted. Judges rated this 192-item pool as follows.

Judging Task

Judges were selected at random from official lists of full-time male undergraduates at Columbia, New York (uptown), Princeton, and Yale Universities. Thirty-six names were drawn at New York University, 84 at each of the others; 288 in all.

A judge evaluated 12 items in each of the 4 categories, 48 items in all. Item order was counterbalanced. Judges were instructed to rate their items on two 7-point scales: *interest* and *ease of refutation*. Verbatim instructions read, in part, "How interested would you be in hearing more about this particular argument" and "How difficult would it be for you to oppose (refute, argue against, counter) this particular argument in a debate. Imagine that you are debating this reason for voting with some other person." Each judge was, in addition, requested to indicate both his political party and candidate preferences. Questionnaires were returned anonymously.

A questionnaire and covering letter were mailed to each judge. The letter appeared to originate with a (fictitious) college debating society which explained that it was involved in a study of factors making for exciting and informative debates, and that the enclosed items had been taken verbatim from recent debates. The judge's cooperation was requested, and he was given detailed instructions and shown a sample of his task. Judges were asked to "try to remain as impartial as possible."

Except for its disguised source the explanation sent to each judge was, in fact, a truthful one in that the process of examining and refuting dissonant information may be considered a sort of intrapersonal "debate." A debating society, then, provided an appropriate setting for judgment of item interest and ease of refutation. Judges were asked their own views and were requested to be impartial in order to reduce the discrepancies in judgments made by persons holding different views. Marked discrepancies in ratings by judges differing in opinion might (but did not) interfere with the selection of items for the letters. Of 288 questionnaires mailed, 10 were undeliverable and 1 was returned blank. Legitimate replies were received from 154 (56% of 278) judges. All but 3 respondents replied to at least 85% of the 96 (48 items × 2 scales) evaluations requested of them; over 99% indicated their political preferences.

The judges' ratings were averaged to obtain interest and ease of refutation scores for each of the 192 items. The following idealized criteria, arranged in descending order of importance, were used in choosing final items from the pool:

1. All items should have equal interest ratings. This criterion attempts to eliminate interest (novelty or information value) as an alternate explanation for the data.

2. All items should have as high interest ratings as possible. This criterion attempts to maximize anticipated low response rates.

3. Half the items should be extremely easy to refute; the others should be as strong as possible.

4. The six items in each message should appear neither redundant nor inconsistent; they should form a natural set. Also, messages should be of equal length.

The 4 letters, each requiring 6 items, led to a selection ratio of 1 in 8 (24 of 192) in choosing items from the judged pool. When the best items were identified, the messages were ready for mailing.

Subjects

Messages were sent to 665 full-time male undergraduates at Brown, Columbia, Princeton, and Yale Universities. Young Democrat and Young Republican Club membership lists yielded 433 (65% of 665) acceptable subjects at Brown, Princeton, and Yale. A telephone poll of students whose names had been chosen at random from registration lists at Columbia, Princeton, and Yale supplied an additional 232 (35%) Johnson Democrats and Goldwater Republicans. (Technicalities prevented the use of political club members at Columbia and polled students at Brown.)

The telephone survey, necessary in order to ascertain the political views of nonclub subjects, obtained political party and candidate preferences. It was carried out by a fictitious political polling organization some 7–14 days before the critical offers were mailed. Interview data later supported our hope that this poll would not be associated with material received in the mails.

The names of club members, polled students, and judges were compared, and persons meeting one or another of the following criteria were eliminated, as specified below, from the subject sample. This procedure was intended to reduce intersubject interaction: (*a*) exclude from sample a judge or a person sharing a room with a judge, (*b*) exclude from poll rosters a person appearing on both club and poll rosters, (*c*) exclude from sample a person appearing on both club rosters, (*d*) randomly exclude all but one of two or more persons sharing a room or apartment.

About 20% of the original sample was thus eliminated, mostly by criterion *d*. This is not surprising for, as the sampling rate within a population increases, the probability that a subject will somehow relate to another rises sharply.

Four types of subjects were drawn in this manner: members of Young Democrat Clubs, members of Young Republican Clubs, polled Johnson Democrats, and polled Goldwater Republicans. These attitude classes will hereafter be referred to as Dems, Reps, J/Ds, and G/Rs, respectively. The final subject pool consisted of 216 Dems, 217 Reps, 169 J/Ds, and 63 G/Rs; 665 in all.[2] Within each school names in each of the four subject classes were assigned randomly but with equal frequency to the four types of messages.

Postexperimental Interview

Letters were mailed to subjects 2 weeks before Election Day. When the flow of replies had all but ceased, interviews were carried out to provide information about (*a*) effectiveness of manipulations, (*b*) reliability of attitudinal information, and (*c*) reasons for reply or nonreply. The interviews were conducted with 138 Princeton and Yale subjects. These schools were chosen since they were conveniently located, and they were the only ones at which both club and polled subjects had been used. The 138 figure represents the number of replying and nonreplying subjects who were reached during 4 evenings of intensive interviewing on location.

All persons who returned postcards or who requested information were informed of the true nature of the affair. A report of the findings was sent to all participants who requested it.

RESULTS

Checks on Message Manipulations

As the messages were composed of items deliberately selected to be of equal interest and of either high or low ease of refutation, it is important to inquire as to the success of the selection procedures. It was assumed that the interest and ease of refutation values of a message would equal the mean interest and ease of refutation values of its component items. Analyses of variance of the judges' ratings of the chosen items indicated no effect for interest ($F = 1.51$, $df = 3/914$) and a very significant effect for ease of refutation ($F = 109$, $p < .01$, $df = 3/912$). These results are not surprising for they merely confirm a selection procedure.

The interviewed subjects were asked to rate their letters on 3- and 7-point ease of refutation scales. Their ratings discriminated significantly between the easy and hard to refute messages (3-point scale analyzed by chi-square, $\chi^2 = 15.8$, $df = 2$, $p < .001$; 7-point scale by analysis of variance, $F = 13.2$, $df = 1/76$, $p < .01$). Finally, if subject cells are collapsed, it can be seen that the response rates to the various messages were not significantly different from each other ($\chi^2 = 1.52$, $df = 3$, $p > .50$).[3] In all, then, the messages appear equally interesting and quite divergent on ease of refutation.

Checks on Subject Attitude Data

Subjects' attitudes were initially ascertained either implicitly (in the case of political club members) or explicitly (when club nonmembers were polled by phone). The postexperimental interviews contained data with which earlier designations could be compared. Over 97% (134/138) of the interviewees verified both their party and candidate preferences. The nature of the shifts in the nonverified cases (e.g., former Goldwater Republican now a Johnson Republican) suggests real uncertainty or attitude change, and not interviewer error, to be the source of unreliability. The identified unreliable cases were excluded from all analyses.

Some General Observations about Postcard Response Rates

Of 665 subjects sent messages, 10 were deleted from the study population as not at indicated address and 4 for attitudinal unreliability. Of 144 replies, 21 were categorized "no reply" because the returned cards strongly suggested a motive other than a desire to obtain a brochure for himself by a

---

[2] The final pool also included 83 polled Johnson Republicans and Goldwater Democrats. The latter subjects are excluded from this report primarily because their divided loyalties hopelessly confound any actor-message relationship.

In addition to the 665 + 83 subjects in Experiment I, 200 others were used for Experiment II. Brown University contributed 230 of the grand total of 948 (24% of the subject population, 9.5% of the Brown undergraduate population); Columbia, 163 (17%/4.6%); Princeton, 236 (25%/7.5%); and Yale, 319 (34%/8.0%).

[3] All probabilities, other than those for analyses of variance, are two-tailed.

TABLE 1

PERCENTAGE OF POSTCARDS RETURNED: PREDICTIONS AND DATA BY SUBJECT AND MESSAGE CLASS

S class	Message class					
	All messages	Strong consonant	Weak consonant	Strong dissonant	Weak dissonant	Mailings per message class[a]
Any (predictions)	—	High	Low	Low	High	—
Dems	21.2	34	11	13	26	53
Reps	18.4	27	15	8	24	53
J/Ds	13.3	15	7	10	21	41
G/Rs	19.7	27	21	6	25	15
Combined sample: Dems, Reps, J/Ds, & G/Rs	18.1	26	12	10	24	160

[a] Plus or minus 2 because the number of subjects available could not always be split into four exactly equal groups and because a few cases were deleted from the initial population (see "Some General Observations about Postcard Response Rates" section).

legitimate subject (13, name and address not that of a legitimate subject; 8, zero brochures ordered). Finally, 5 other cards were categorized 'no reply" because they originated after interviewing was in progress. The analyses are based on 118 replies from 651 mailings.

Club members were, on the whole, slightly more receptive to the offer than were nonmembers. Dems and Reps returned 19.8% of their cards, versus 15.0% for J/Ds and G/Rs ($\chi^2 = 2.02$, $df = 1$, *ns*). This effect was contributed almost entirely by Johnson supporters. Dems returned 21.2% of all their cards, whereas J/Ds returned only 13.3% ($p < .05$ by normal approximation to binomial). Reps returned 18.4% and G/Rs, 19.7% (difference not significant).

The data were collapsed across schools, as between-school analyses of overall response rates proved nonsignificant (Experiment I: $\chi^2 = 2.80$, $df = 3$; Experiment II: $\chi^2 = 1.75$, $df = 2$).

Evaluation of Hypotheses

Table 1 gives the return rates for each of the four messages by subject attitude class. It is obvious from Table 1 that the data support the predictions. In each of the four subject classes strong consonant and weak dissonant messages are preferred to weak consonant or strong dissonant. The statistical significance of the effects can be tested with chi-square comparisons. For example, the significance of the preference for strong consonant over weak consonant information by Dems is evaluated by comparing the 18 returns from 53 mailings (34%) for strong consonant versus 6 returns from 54 mailings (11%) for weak consonant. Unfortunately, the breakdown into subject classes results in cells whose *N*s are so meager as to preclude classical levels of acceptable significance. In order that weak positive findings be discriminated from nonsignificant ones, all probabilities of 1 in 4 or better will be so indicated for subgroups of the total population.

TABLE 2

STRENGTH OF PREDICTED EFFECTS: EXPERIMENT I

*S*s	Hypothesis I: Strong consonant preferred to weak consonant	Hypothesis II: Weak dissonant preferred to strong dissonant	Hypothesis III: Strong consonant preferred to strong dissonant	Hypothesis IV: Weak dissonant preferred to weak consonant
Dems	<.01	<.20	<.025	<.10
Reps	<.25	<.05	<.02	*ns*
J/Ds	*ns*	<.25	*ns*	<.20
G/Rs	*ns*	*ns*	*ns*	*ns*
Combined sample: Dems, Reps, J/Ds, & G/Rs	<.005	=.001	<.001	<.01

Note.—All nonsignificant differences are in the predicted direction. All analyses by chi-square.

The chi-square probabilities are given in Table 2. Although the significance levels for the individual subject classes are often marginal, the overall pattern of results is encouraging. Without exception, all the data fall in the correct direction. And in the combined sample, the differences are highly significant. The findings clearly support the model.

Postexperimental Interviews

Early in the interviewing schedule, respondents were asked two open-ended questions about recent exposure to political materials. We were interested in observing recall of the mailed letters. Only 41 of 138 subjects (30%) spontaneously recalled the letter they had received but 6–9 days earlier. One is struck by the number of subjects who were unable to recall the message they had received even when handled a bundle of the letters used and asked to identify the one they had been sent. Of 138, 9 simply did not recall any letter; 21 others did not correctly identify theirs even after twice examing the package. Six other subjects, while correctly identifying the message, erred in recalling whether or not they had returned the postcard. In all, 36 of 138 (26%) did not correctly reidentify these events which had, presumably, engaged their attention some short time ago. The 26% figure should be further increased (say, to 33%) since there were doubtless other persons who could not remember but who guessed correctly. Perhaps the letters had never been considered with the care and attention one reserves for "truly important matters" (i.e., experiments). We will return to this issue later.

When asked why they had or had not returned the postcards, subjects in all experimental cells tendered similar and bland explanations. Those who had returned their cards usually explained that they wanted to be better informed and that they were curious about the facts or about the organization. Those who did not return their cards explained that the arguments were petty, that there was nothing to be learned, that they were already well informed, that they were just not interested. These bland explanations, *invariant across experimental conditions,* do not account for the sharp differences in response rates to the different messages. They strike one rather as reasons the respondent makes up when, prodded by the interviewer, he puzzles over his own unconsidered previous actions. Here too we would suggest that the behavior the data strongly suggest did occur was, on the part of the actors, not a product of deliberation.

EXPERIMENT II

Experiment II also offered subjects politically relevant brochures. However, all subjects in the second study received the same offer. That offer appeared to originate with a nonpartisan source which was interested in helping voters but not in supporting any particular candidate. The organization displayed samples of a variety of brochures which might be ordered, free of charge, by returning a postcard. The subject could order any four of the eight brochures displayed. The eight brochure samples systematically filled a 2 × 2 × 2 design: Candidate × Ease of Message Refutation × Positive or Negative Argument. Who, we again inquired, sends for what?

Methodological Details

The eight critical stimuli (brochure samples) were drawn from the 24 stimuli used in the first study. Analyses of variance of the interest and ease of refutation ratings of the eight stimuli again supported the belief that they were of similar interest ($F < 1$) and very different on ease of refutation ($F = 12.7$, $p < .01$, $df = 7/294$).

The nonpartisan offer was mailed to 100 Dems (31% of the Dems in each school) and 100 Reps (32%). Of course, these subjects did not participate in Experiment I.

In most respects the two experiments were quite alike. The offering letters were similarly structured. Both studies shared one stimulus selection procedure and one subject selection procedure. Both relied on postcard return data. Both were carried out at the same time. Experiments I and II differed, however, in the following: I utilized a partisan source; II, a nonpartisan source. I offered a choice of a specific brochure or nothing; II, a forced choice of 4 of 8 brochures. Each brochure in I was represented by six arguments; in II each was represented by only one argument. Our initial considerations did not lead us to expect qualitative differences between the two experiments.

RESULTS

Brochure requests were returned by 19 Dems and 40 Reps. The 19% Dems return

TABLE 3

NUMBER OF SUBJECTS IN EXPERIMENT II WHOSE DATA CONFIRM OR REJECT THE HYPOTHESES

Ss' attitude	No. *Ss*			
	Hypothesis I: Strong consonant preferred to weak consonant	Hypothesis II: Weak dissonant preferred to strong dissonant	Hypothesis III: Strong consonant preferred to strong dissonant	Hypothesis IV: Weak dissonant preferred to weak consonant
Dems ($N = 19$)				
Confirm	5	4	5	7
Reject	4	6	6	7
Not assessable	10	9	8	5
Reps ($N = 40$)				
Confirm	20[a]	16	30[b]	10
Reject	5	11	4	18
Not assessable	15	13	6	12

[a] 20 versus 5 significant at $p < .01$ by chi-square.
[b] 30 versus 4 significant at $p < .001$ by chi-square.

rate was not much different from the 21% Dems return rate of the earlier study. The 40% Reps rate was, however, significantly higher than the 18% Reps rate in Experiment I ($p < .01$ by normal approximation to binomial). Perhaps the 40% return rate resulted from a relative scarcity, at our sites, of highly credible source material supporting Goldwater.

Each of the four hypotheses was evaluated against the preference pattern of each responding subject. Nonresponders were thus ignored in Experiment II (but not in I). Hypothesis I, for instance, expects strong consonant to be preferred to weak consonant. A subject could request zero, one, or two brochures of any of the four message types offered (strong consonant and others), with a total of four requests among the eight available brochures. If a subject requested two strong consonant brochures and only one weak consonant, then his choice pattern confirmed the hypothesis. As he had only four choices, one strong consonant request and one weak consonant request by a subject put him into a "nonassessable" category vis-à-vis Hypothesis I. If he requested one strong consonant and two weak consonant brochures, his data rejected Hypothesis I. If a hypothesis is valid, then significantly more subjects will confirm it than will reject it. If it is not valid, then the number of subjects confirming it should not exceed the number rejecting it. The results are shown in Table 3.

It is clear from Table 3 that the few Dems exhibited no selective preferences in this study. They were as likely to request one brochure as another. The Reps, however, preferred strong consonant to weak consonant (Hypothesis I, $p < .01$, Table 3), weak dissonant to strong dissonant (Hypothesis II, *ns*), strong consonant to strong dissonant (Hypothesis III, $p < .001$, Table 3), and weak consonant to weak dissonant (Hypothesis IV *upset, ns*). The latter finding is disturbing because it fails to confirm the central result of the earlier study. The following section will consider this problem in some detail.

DISCUSSION

The results of Experiment I clearly support the hypotheses about the relative desirability of strong consonant and weak dissonant over weak consonant and strong dissonant information. The predicted effects emerge in all groups and achieve high levels of significance in total sample comparisons. The finding that weak dissonant information is preferred to weak consonant information refutes the alternative view that weak messages are merely not avoided (rather than sought out).

The results of Experiment II are not as easily interpreted. While the Reps showed preferences among messages, the Dems did not. The Reps' preferences supported all the hypotheses, save the critical one which dealt specifically with the preference for weak dissonant over weak consonant information. A number of explanations suggest themselves, any one of which may account for the latter negative finding without disturbing the broader theoretic model. Three are given below.

1. Ease of argument refutation is likely related to message source. Fallible dissonant arguments are probably perceived weaker if linked to a partisan than to a nonpartisan source. Manis (1961) found that dissonant arguments linked to a high prestige source were perceived closer to one's own opinion than were the same arguments linked to a low prestige source. The items used in the nonpartisan source study (Experiment II), while consisting of some of the very same items used in the partisan source study (Experiment I), may not have been perceived as sufficiently vulnerable to merit examination. It is not clear whether dissonant messages intermediate in ease of refutation will be approached or avoided. According to the model there must exist a critical intermediate value of ease of refutation at which the approach and avoidance tendencies are balanced. Only when one could identify that critical value could one make predictions for all messages. Since the weak messages in Experiment II (nonpartisan source) were probably perceived as less weak than the same arguments in Experiment I (partisan source), it is unclear as to whether we should expect Hypothesis III or Hypothesis IV to apply.

2. Experiment I used six arguments for each message. Experiment II used but one per message. Perhaps the intended impression of message ease of refutation was more successfully communicated in Experiment I than in Experiment II.

3. Environmental demands for rationality (one should be open to all points of view) probably affect selective exposure. For this very reason subjects in the present studies were unaware of experimental intent in their mail. Nevertheless, a nonpartisan source, offering a forced-choice among booklets, may provide subtle cues which elicit a rational posture. By employing a blatantly partisan source in the first study, and especially by offering the subject only one brochure, we may have succeeded in catching the subjects off their intellectual guard. The large number of subjects in Experiment I who were unable to remember either the letter or their reaction suggests the superficial level at which the offer was handled.

Because of the above systematic deficiencies in the second study it cannot strictly be considered an unsuccessful replication of the first. The consistency and overall significance of the effects observed in the first study suggest that, in at least some settings, people do prefer dissonant to consonant information. To be sure, however, this finding remains in need of replication.

It is appropriate to cite other reported instances of deliberate exposure to dissonant information. Brock (1965) and Freedman (1965b) suggested deliberate exposure to dissonant information as an explanation for some of the negative findings in recent selective-exposure research. In a report by Canon (1964) high- but not low-confident subjects preferred dissonant to consonant messages.[4] A similar interpretation might be placed on two studies reported by Brehm and Cohen (1962, Ch. 6) in which dissonant information was recalled better than consonant. Even more interesting is a study by Jones and Rika (1958) in which subjects remembered plausible consonant and implausible dissonant arguments better than implausible consonant or plausible dissonant ones. While the Jones and Rika data may be due to selective forgetting in the service of consistency, they may also be a result of differential attention paid the various arguments. Finally, small-group communication and conformity studies by Brodbeck (1956), Festinger and Thibaut (1951), Gerard (1953), and Schachter (1951) showed that, in discusison situations, people preferred communication with disagreers rather than with agreers. In maintaining cognitive consistency, man's tactics appear far from simpleminded.

---

[4] Freedman (1965a) cannot replicate the Canon finding.

REFERENCES

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance*. New York: Wiley, 1962.

BROCK, T. C. Commitment to exposure as a determinant of information receptivity. *Journal of Personality and Social Psychology,* 1965, 2, 10–19.

BRODBECK, M. The role of small groups in mediating the effects of propaganda. *Journal of Abnormal and Social Psychology*, 1956, **52**, 166–170.

CANON, L. K. Self-confidence and selective exposure to information. In L. Festinger, *Conflict, decision, and dissonance*. Stanford: Stanford University Press, 1964. Pp. 83–95.

FESTINGER, L. A theory of social comparison processes. *Human Relations*, 1954, **7**, 117–140.

FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *Journal of Abnormal and Social Psychology*, 1951, **46**, 92–100.

FREEDMAN, J. L. Confidence, utility, and selective exposure: A partial replication. *Journal of Personality and Social Psychology*, 1965, **2**, 778–780. (a)

FREEDMAN, J. L. Preference for dissonant information. *Journal of Personality and Social Psychology*, 1965, **2**, 287–289. (b)

FREEDMAN, J. L., & SEARS, D. O. Selective exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965. Pp. 57–97.

GERARD, H. B. The effect of different dimensions of disagreement on the communication process in small groups. *Human Relations*, 1953, **6**, 249–272.

JONES, E. E., & RIKA, K. The effects of plausibility on the learning of controversial statements. *Journal of Abnormal and Social Psychology*, 1958, **57**, 315–320.

LOWIN, A. Information selectivity as a function of agreement with message and ease of message refutation. Unpublished doctoral dissertation, Columbia University, 1965.

MANIS, M. The interpretation of opinion statements as a function of recipient attitude and source prestige. *Journal of Abnormal and Social Psychology*, 1961, **63**, 82–86.

SCHACHTER, S. Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, 1951, **46**, 190–207.

(Received March 9, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 10–15

DEFENSIVENESS IN DECISION MAKING AS A FUNCTION OF SEX-ROLE IDENTIFICATION[1]

PAUL D. LIPSITT AND FRED L. STRODTBECK[2]

University of Chicago

Styles of defenses invoked by individuals representing varying sex-role identities were distinguished in a decision-making task. Male Ss who were unconsciously feminine and consciously masculine were more likely to find a defendant guilty in a simulated trial when homosexuality is ascribed to the defendant. Male Ss who were both unconsciously and consciously feminine were more likely to find a defendant not guilty when the question of moral character is combined with the homosexual reference. Males who were both consciously and unconsciously masculine vary least in their decisions in the several versions of the trial.

In *Inner Conflict and Defense* (Miller & Swanson, 1960), studies are reported which explore the relationship between an individual's sex-role identity and his responses when stimuli selected to arouse sexual or aggressive drives are introduced into the experimental situation. In these studies subjects were classified into sex-role identity types, in terms of their scores on two tests of masculinity-femininity—the Franck Drawing Completion Test (Franck, 1949) and the Gough Femininity Scale (Gough, 1952). Despite the fact that both the Franck and Gough tests were standardized on their ability to discriminate boys from girls, a nonsignificant correlation between the scores on these two tests has repeatedly been demonstrated (Cochrane, Strodtbeck, & Parkman, 1965; Miller & Swanson, 1960; Shepler, 1951). Scores on the two tests can therefore be used simultaneously to classify subjects into one of four types—masculine-masculine (MM), masculine-feminine (MF), feminine-masculine (FM), and feminine-feminine (FF). By convention, scores on the Franck test are listed first.

Lansky (1960, Ch. 12) utilized this composite measure to investigate the behavior of subjects of varying sex-identity type under conditions of sexual arousal. It was found that FF males show a significantly greater tendency, after sexual arousal, to write endings exhibiting denial—scored if an ending was obviously incompatible with one of the facts in the story beginning; or withdrawal—scored if the ending described the hero as having removed himself from the scene of the conflict. FM males show a significantly greater tendency to write endings in which the hero expresses aggression against self—scored if the hero ignored the authority's wise counsel and suffered the consequences or if the ending described an injury to the hero. Of the three sex-role categories in Lansky's data, the MMs had the highest probability of realistic problem solving after sexual arousal.

Miller and Swanson used Lansky's study to support their formulation of relationship between sex-role identity and selection of defenses. These authors view the Franck Drawing Completion Test as a measure of "unconscious" masculinity-femininity, in that scores on the Franck test are presumed to reflect an underlying identification with the male or female role, perhaps based upon early identification with the father or the mother. Scores on the Gough test are presumed to reflect "conscious" masculinity-femininity, or the degree of overt acceptance of preferences, attitudes, and interests which are culturally defined as appropriate for males.

---

[1] This article is based on a doctoral dissertation by the senior author submitted to the Department of Psychology, Division of Social Sciences, University of Chicago. The Russell Sage Foundation supported a residency for the senior author during his graduate training.

[2] The senior author is now at the Law-Medicine Institute, Boston University. The senior author is indebted to David Bakan and Eugene Southwell for their guidance as members of his dissertation committee. Further thanks are due to Judith Fritz for scoring the data and to Margaret Parkman for her critical reading of an earlier manuscript.

The present study utilizes the Franck-Gough composite measure of sex-role identity and is an extension of Lansky's research on the behavior of subjects of varying sex-identity type under conditions of sexual arousal. Rather than employing a traditional projective test to elicit responses, in the present study a decision-making task was presented in a simulated institutional context (a legal trial) to determine whether the variations in response will be comparable to those found under projective conditions. The task orientation of rendering a verdict in a legal trial provided the subjects with a set which is different, and perhaps more pervasive, than the traditional projective technique. Typically, story-telling or story-completion devices provide data for personality analysis. In contrast with this kind of fantasy, the trial provides a simulated institutional context within which to test behavioral responses. If sex-role identity is as differentiating and pervasive a personality dimension as Miller and Swanson imply, then the decision-making process employed in the institutionalized context of a jury trial should also be different for individuals in the various sex-role categories.

METHOD

Experimental Task

The trial involved a former American soldier accused of treason as a result of alleged activities prejudicial to the United States while a Japanese prisoner of war after the fall of Corregidor in 1942. The trial judge had allowed the introduction of testimony of the defendant's alleged homosexuality on the grounds that it went to the character of the defendant and was therefore admissible. On appeal, it was held that this testimony, elicited as part of the evidence in the trial, was irrelevant, inflammatory, and prejudicial to the issue at bar. The experimental recreation of the trial was designed to control for both admission of homosexuality into the testimony as well as reference to the defendant's moral character. These controls determined whether the "arousal materials" had an overall influence upon the decision and whether the influence differentiated subject jurors as a function of their sex-role identity.

Subjects

The subjects were 380 male naval enlisted personnel between the ages of 18 and 29 who were assigned to a hospital corpsman training school.

Treatment Variables

Four versons of the case were developed, using the various combinations of two types of arousal materials as the only variations in the tape recording of the testimony, which otherwise remained the same under the four treatments. The treatment variables, determined by the arousal materials, provided a 2×2 design, based on the presence or absence of one or both types of arousal material.

One type of arousal material dealt with the suggestion of homosexuality in the defendant, while the other type of arousal material was based on suggestions by the judge in his charge to the jury relative to the defendant's good or bad character. The jury was instructed about the permissibility of such evidence in arriving at their finding of *guilty* or *not guilty*. When homosexual references to the defendant were absent, it was designated as h and when present, as H. The judge's intervention when absent in a treatment variable was referred to as a, and when present, as A.

The basic case (ah) contained neither of the two types of arousal material. All evidence in this version of the case was present in the other three versions of the trial. Only the specific arousal materials were added in the variations from the basic trial.

Introduced in the testimony in H was the cross-examination of a defense witness concerning the interest of the defendant in doing the laundry for a colonel. Upon ascertaining that the defendant had done a specific colonel's laundry, the prosecuting attorney then asked, "And do you know whether this is the same Colonel Crane who later received a discharge from the Army as a result of evidence of homosexuality?" Another item in the testimony of the H arousal material dealt with the cross-examination of a neuropsychiatrist in whose division the defendant had been a patient. The critical question on cross-examination that opened the direct accusation of homosexuality was, "And do these records state that in November, 1946, Renfrew (the defendant) was hospitalized in Camp Lee, Virginia, because of homosexual aberrations?" Over strenuous objection by defense counsel, the witness was allowed to testify that there were indications that the defendant had been involved in some discussions with another patient that may have been interpreted as homosexual in nature. The cross-examination went on to indicate that Renfrew was incarcerated because of an obscene conversation concerning homosexual aberrations between himself and another soldier. In the prosecutor's summation to the jury in H, he was permitted to say, "I will confine myself to the facts and will not elaborate on any odious sexual behavior or deficiencies of moral character that may exist in the defendant."

In A, the judge, in charging the jury, added his opinion to the standard instructions of a:

> Evidence of the defendant's good character is in the same category as other factual evidence, and must be considered by you in your deliberations. Such evidence of the defendant's good or bad character may, of itself, if believed by you,

create a reasonable doubt, where otherwise no reasonable doubt would exist. If you believe that the behavior of the defendant was indicative of odious moral character, then that too can be taken into consideration in determining whether or not he is guilty of the offense charged. He is being tried for treason, but all of the evidence of the defendant's character is available to you in reaching your decision.

Personality Variables

Sex-role identity. Sex-role identity is used to refer directly to the classification of subjects by simultaneous consideration of two tests: a projective technique based on drawing completions (Franck, 1949), and a true-false Likert-type attitude scale based on self-report (Gough, 1952).

The Franck stimuli, each consisting of a few simple lines, were presented to the subjects in a booklet containing one drawing on a page, with the following instruction:

> In the following pages, you will find a number of incomplete drawings; please complete them; do it any way you like; use as many lines as you wish; do it the way it seems the most fun. There is no right or wrong way of doing this.

The scoring was based on six formal criteria or principles and a content category. Each drawing received a resultant score of Masculine (M) or Feminine (F). The summing of Fs determined the total score, and therefore the higher the score, the more unconscious femininity. A scoring key[3] indicated which criteria were to be used for the different stimuli and provided rules for coding complex responses. For example, expanded drawings were scored "male," while drawings left open were scored "female." Active objects, such as motorboats or steamboats, were "male" and passive objects like sailboats and containers, "female." Ten of the original 36 Franck forms were used. These 10 forms were found by Cochrane et al. (1965) to discriminate best between males and females of the entire set in the original instrument. A reliability of .91 between scorers was found for a sample of data in this study.

The second measure of sex-role identity was obtained from the Gough Femininity scale. This instrument consists of 58 items eliciting "true" or "false" reponse to such items as "I want to be an important person in the community"; "I prefer a shower to a bath"; and "I would like the work of a dress designer." By dichotomizing the scores for each test at the mean, four categories of subjects were identified.

Procedure

During the first week of their training period, hospital corpsmen comprising a company or class were administered the tests which formed the basis for the personality variables in the study. While seated in a classroom, the subjects were given the following instruction:

> My name is ____________, and I'm from the University of Chicago, where we are conducting research. In line with the purposes of this research, we are collecting data from you. This information will be confidential and will in no way be a part of your service record. The information is to help us to understand more about groups of people and will not be used to assess you individually.

The subjects were then provided with a packet containing the short form of the Franck Picture Completion Test. Immediately after finishing the Franck test, the subjects responded to the items on the Gough Femininity scale.

Eight weeks after the administration of this test material, the subjects were presented with one of the four versions of the trial described above. The subjects, who heard the taped testimony seated together as a class group, were read the following introductory statement:

> The voices you will hear are not the principals in this case. Our recording has been prepared from the transcript of the proceedings; the names of the principals have been changed to protect their identity and repetitious sections of the testimony have been condensed, but aside from these exceptions, you will hear a re-creation of what took place in court. While the case is now closed, and your actions can in no way affect the principals, we wish you to adopt the attitude of responsibility and duty which you would bring as citizens serving on a regular jury. After the trial you will be asked to decide the case. Our recording begins when the judge is directing the prosecution to open its case.

The testimony of the 1-hour trial was then presented by tape recorder, with a 5-minute recess between the case for the prosecution and that for the defense. Immediately after hearing the testimony of the trial, each subject received a ballot on which to indicate whether he thought the defendant was *guilty* or *not guilty* of the offense of treason.

RESULTS

In the basic version (ah) and the version containing only the judge's intervention (Ah), the observed frequencies were not significantly different among the sex-role categories from what might be expected by chance. In the ah version, it was hypothesized that there would be no differential arousal since this version of the case contained none of the materials that were hypothesized to be differentially stimulating

---

[3] Copies of this manual were obtained from Leonard M. Lansky, University of Cincinnati. He obtained the manual from S. Dunn, Australian Council for Educational Research, Melbourne, who holds the copyright for the test.

TABLE 1

CHI-SQUARE ANALYSIS OF DIFFERENCES ACROSS SEX-ROLE IDENTITY CATEGORIES FOR EACH TREATMENT OF THE CASE

Treatment	Guilty			Not guilty			Total χ^2
	Observed	Expected by chance	χ^2	Observed	Expected by chance	χ^2	
ah							
MM	16	17.08	.068	16	14.92	.078	.382
MF	19	17.62	.108	14	15.38	.124	
FM	14	13.33	.001	12	12.12	.001	
FF	14	14.42	.001	13	12.58	.001	
Ah							
MM	9	9.63	.087	6	5.07	.171	.508
MF	10	9.93	.003	5	5.07	.000	
FM	17	15.89	.078	7	8.11	.152	
FF	11	11.25	.006	6	5.75	.011	
aH							
MM	21	23.08	.188	16	13.91	.311	10.043*
MF	15	18.09	.528	14	19.90	.881	
FM	26	18.72	2.792	4	11.28	4.698	
FF	16	18.09	.241	13	10.90	.404	
AH							
MM	11	10.64	.012	7	7.36	.018	10.272*
MF	10	7.09	1.194	2	4.91	1.725	
FM	10	7.63	.701	3	5.32	1.012	
FF	8	13.59	2.299	15	9.41	3.321	

* $p < .02$.

for the sex-role identity types. In the Ah version, with the introduction of the judge's charge regarding the moral character of the defendant, there was an increased tendency by all personality types to find the defendant *guilty*. Table 1 shows that among the four sex-role categories the tendency to find the defendant *guilty* was not significantly different in this version of the case. In the absence of the implication of homosexuality, this intervention of the judge seemed to influence *guilty* findings uniformly in all subjects, as indicated in the Ah version of the case. A chi-square of .508 does not demonstrate significant differences among the four sex-role categories.

In the aH version of the case, the combined chi-square for the four sex-role categories indicated that the distribution does not follow a chance pattern ($\chi^2 = 10.043$, $df = 3$, $p < .02$). First, chi-square comparisons between the FM group and each of the other groups indicated that the frequency of *guilty* responses was significantly greater than that of any other group with the introduction of homosexuality into the case ($\chi^2_{\text{FM, FF}} = 5.68$, $p < .02$; $\chi^2_{\text{FM, MF}} = 6.92$, $p < .01$; $\chi^2_{\text{FM, MM}} = 5.72$, $p < .02$).[4] The very high percentage of *guilty* findings by FMs, 87%, lends support to the proposition that the threat-inducing arousal materials caused a distortion of the basic issue of treason.

The AH version of the case, with the combination of homosexuality and intervention by the judge, showed an overall difference in frequencies of *guilty* findings among the four sex-role categories. A combined chi-square of 10.272 attained significance at the .02 level. Most of the variation in these decisions is attributable to the FFs, who responded to the combination of homosexual references in the trial and the judge's intervention with a significant shift toward *not guilty* findings. Chi-square analysis of the FFs with each of the sex-role categories indicated that the FFs had been differentially affected by the combination of the judge's charge and the homosexual references ($\chi^2_{\text{FF, FM}} = 4.33$, $p < .05$; $\chi^2_{\text{FF, MF}} = 5.72$, $p < .02$; $\chi^2_{\text{FF, MM}} = 8.59$, $p < .01$).

---

[4] The Yates correction for continuity was applied in all comparisons between specific sex-role categories in order to accommodate small cell frequencies (Guilford, 1956, p. 234).

DISCUSSION

Based upon an analysis of their responses to the trial, the MMs were found to be the most reality oriented of the four sex-role categories. That is, they were minimally affected by the arousal materials, varying least from the 50% base line of *guilty* findings in all versions of the case. In addition, in the basic version, the other three sex-role categories did not differ significantly in *guilty* findings from the MMs, suggesting that the impact of the basic version of the trial was not differentially arousing to the various sex-role categories.

That the MMs would respond more similarly in all four versions of the case is consistent with the findings of Lansky that the MMs are least defensive when confronted with sex arousal material. In the task involved in this study, the MMs relied fundamentally upon the issue of treason, the offense for which the defendant was being tried, and were least influenced by the introduction of the prejudicial materials which were not directly relevant to the legal issue under consideration.

The MFs did not contribute to any significant differences in responses to the various versions of the trial. Generally, the responses of the MFs were similar to those of the MMs. The lesser change in the face of extralegal considerations of both the MM and MF males suggests that males with masculine scores on the projective drawing-completion test have developed masculine behavioral dispositions of a sufficiently basic and pervasive character to be secure in their sexual identity and thus relatively free of disruptive anxiety when evaluating sexually deviant behavior. Indeed, one might speculate that masculine scores on the Franck test reflect behavioral dispositions which generate such a consistently masculine impression that the MF male is less subject to social pressures. His interests and preferences, reflected in his feminine responses to the Gough test, can be expressed without regard for the fact that they are culturally stereotyped as feminine.

With the introduction into the trial of a suggestion of homosexuality, a significant majority of the FM subjects found the defendant *guilty*. The percentage of FMs returning verdicts of *guilty* was significantly higher than that of any other sex-role category in the aH version of the case. Our understanding of the FM's punitive response toward the defendant in the aH version is aided by reference to Miller and Swanson's description of the FM as "at war with himself." Let us assume that the FM's feminine scores on the Franck test reflect basic personality dispositions to behave in a feminine manner, while his masculine scores on the Gough test reflect an acceptance of socially defined sex-role attributes, attitudes, and preferences as appropriately characteristic of the self. If such is the case, the FM must frequently have experienced strong tendencies or desires to behave in ways society determined as feminine and therefore undesirable. Having accepted society's sex-role definition, he must also have frequently censored these tendencies in himself. The FM therefore might be seen as a man accustomed to punishing socially deviant behavior of a sexual character. Having learned to be self-punitive toward his own feminine tendencies, he will not hesitate to be punitive toward the defendant for his homosexual tendencies.

The empirical data are not as clear in explaining the impact of authority upon the FMs. While the intervention of the judge increased guilty findings among the FMs, the AH version did not produce a cumulative effect, as might be predicted. The FMs have been characterized as accepting social pressure. If the moralizing of the judge is interpreted as a form of social pressure, then in combination with the homosexual references it might be expected that the AH version would produce an even higher percentage of *guilty* findings than the aH version.

It is notable that the arousal material of homosexuality to which the FMs are so responsive failed to elicit responses in the FFs that significantly deviated from the base-line percentages. Only 55% of the FFs found the defendant *guilty* when homosexual references alone were introduced in the trial. While these two sex-role categories are similar in their disposition to behave in a feminine manner, they differ in their acceptance of cultural stereotypes of appropriate masculine behavior. The FFs are not motivated to reject all traits that are not masculine, as reflected in their scores on the Gough Femininity scale. They are capable of accepting their own femininity, and thus are not aroused by the introduction of the homosexual material in the trial.

The combined version of the trial resulted in a marked change in the FFs' responses. These subjects were unique among the four sex-role categories in reducing the percentage of *guilty* findings to 35% when both arousal materials were introduced. This is the only instance where the percentage of *guilty* findings fell below 50% in any version of the case. The FFs ignored the homosexual references in the trial until the judge intervened with what the FFs apparently perceived as an attack upon the defendant, not in terms of the alleged crime of treason, but in terms of a personal morality that related to the suggestion of homosexuality. Since 65% of the FFs found the defendant *guilty* when the intervention of the trial judge was present alone, it is clear that the interaction of the two types of arousal materials caused the differential response pattern in the FFs. The FFs, in this version of the case, allowed the tangential issue of homosexuality to direct them toward a *not guilty* finding. They overreacted to a sense of "injustice" on the grounds of a peripheral impugnment of the defendant's character.

Lansky has postulated the mechanism of denial to explain the defensive style invoked by the FFs. This rationale might be accepted in this study to explain the "denial" of guilt by the FFs in the AH version. The apparent lack of reaction to the homosexual material by the FFs (55% guilty findings in the aH version) leads to the conclusion that the moral charge by the judge in combination with the homosexual references results in a rejection of *guilty* findings.

The interaction between the homosexual evidence and the presence of the judge's instructions seems inconsistent with the Miller and Swanson formulation. If males who earn feminine scores on the Franck test are "unconsciously feminine" in the psychoanalytic sense of "beneath the layer of repression," and if unconsciously feminine males invoked defense mechanisms to manage the threat of the homosexual suggestion, then the strength of these defense mechanisms should not be as strongly influenced by other variations in the trial stimulus as it in fact is. While this study may not be a crucial test of the Miller and Swanson theory, it does cast doubt upon the defensive formulation.

REFERENCES

COCHRANE, C. T., STRODTBECK, F. L., & PARKMAN, M. *A masculinity-femininity measure to predict defensive behavior.* Chicago: University of Chicago, Social Psychology Laboratory, 1965. (Mimeo)

FRANCK, K., & ROSEN, E. A projective test of masculinity-femininity. *Journal of Consulting Psychology,* 1949, 13, 247–256.

GOUGH, H. G. Identifying psychological femininity. *Educational and Psychological Measurement,* 1952, 12, 427–439.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill, 1956.

LANSKY, L. M. Mechanisms of defense: Sex identity and defenses against aggression. In D. R. Miller & G. E. Swanson (Eds.), *Inner conflict and defense.* New York: Holt, 1960. Pp. 272–288.

MILLER, D. R., & SWANSON, G. E. *Inner conflict and defense.* New York: Holt, 1960.

SHEPLER, B. A comparison of masculinity-femininity measures. *Journal of Consulting Psychology,* 1951, 15, 484–486.

(Received April 8, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 16–22

INTERACTIVE EFFECTS OF SELF-ESTEEM AND TASK DIFFICULTY ON SOCIAL CONFORMITY[1]

KENNETH J. GERGEN AND RAYMOND A. BAUER
Harvard University

While a linear relationship between self-esteem and susceptibility to social influence has often been found among male populations, this relationship has not generally been found among females. The present study was based on earlier indications that the relationship in female populations might be curvilinear in nature. It was further predicted that the curvilinear relationship would occur under task conditions in which judgments were of moderate difficulty. Under both high- and low-difficulty task conditions, it was hypothesized that conformity in general would be less and unrelated to self-esteem level. Evidence from 2 studies, in which 67 females were influenced to change their judgments of artworks, largely supported the theoretical speculations.

One of the more stable findings relating personality and social influence is a linear relationship between level of self-esteem and conformity. Janis (1954), Berkowitz and Lundy (1957), Janis and Field (1959), Lesser and Abelson (1959), and Linton and Graham (1959) have all found that persons low in self-esteem are more persuasible than those whose self-regard is more substantial. However, such findings have been almost totally limited to male populations. When female populations have been used, significant correlations between self-esteem and persuasibility have been virtually nonexistent. Until a recent study by Cox and Bauer (1964), there was little to explain this enigma. These investigators unexpectedly found a strong curvilinear relationship between self-esteem and persuasibility.[2] Contrary to low-self-esteem males who seem to seek social approval through acquiescence, low-self-esteem females were interpersonally defensive and rigid when confronted with opinions differing from their own. In essence, the syndrome of low self-esteem for females may differ from that of males. The present experiment is an attempt to overcome some of the methodological problems inherent in the earlier work, and to test the relationship between self-esteem and persuasibility for females in a controlled laboratory setting.

It also seems useful to consider the range of situations in which a curvilinear relationship is most likely to be found. At least one relevant set of variables has to do with the nature of the stimulus about which judgments are formed and/or altered. For example, responsiveness to social influence has commonly been found to be a linear function of task complexity (Asch, 1958; Blake, Helson, & Mouton, 1956; Coleman, Blake, & Mouton, 1958; Hochbaum, 1954; London & Lim, 1964). The general rationale for such findings is contained in Hochbaum's (1954) statement that "a person's dependence on social referents is inversely related to his confidence in his competence to judge the issue in question [p. 679]." In this light, how might females differing in self-esteem react to social influence under tasks varying in complexity? Data from the Cox and Bauer (1964) study suggested that under simple task conditions persons may be sufficiently confident in their judgments that personal styles may play little role. However, as task difficulty increases persons may become increasingly sensitive to the opinions of others; personality styles should begin to function, and level of self-esteem should begin to exert its effects.

The discussion thus far may be summarized in the form of two hypotheses upon which the present study attempted to shed light:

1. While under simple task conditions female subjects will not differ in conformity as

---

[1] This research was supported by funds from the National Science Foundation (GS-562). We are indebted to Christina Platt and Julia Spring for their services in carrying out experimental procedures.

[2] This finding was also supported in a study of reactions to television commercials, reported in an unpublished manuscript by J. A. Barach, at the Harvard School of Business Administration.

a function of self-esteem level, under conditions of moderate task difficulty, the relationship between self-esteem and conformity will be curvilinear. Females in the medium-self-esteem range will be most persuasible.

2. At all levels of self-esteem, conformity will be greater under a moderately difficult than an easy task.

As will be noted, the above hypotheses are limited to simple and moderately difficult task conditions. During the execution of the experiment designed to test these hypotheses, the question arose as to whether the curvilinear relationship would hold if task difficulty were indefinitely increased. The literature on task difficulty and persuasibility strongly suggests that *any* increment in task difficulty will produce a corresponding increase in the person's reliance on others for information. On the other hand, it seems that if a person were engaged in a virtually insoluble task, another's opinions would add little in the way of useful information, and sensitivity to others would thereby lose its functional value. In essence, as a task becomes exceedingly difficult, one might expect a shift in problem-solving strategy away from depending on others for useful information (cf. Bruner, Goodnow, & Austin, 1957). For present purposes, two results might be expected from such a shift in strategy. First, rather than the linear relationship so often suggested, if a task is made exceedingly complex one might predict the development of a curvilinear relationship between social conformity and task difficulty. Under moderately difficult task conditions, the greatest conformity would be evidenced. The second implication relates to the effects of self-esteem. Similar to simple task conditions, when a task becomes inordinately difficult and decisions become arbitrary, persons should again become immune to personal influence. Individual differences in self-esteem should again fail to predict amount of conformity. This latter discussion can be summarized in the form of two additional hypotheses:

3. Conformity will be less under an extremely difficult task than a moderately difficult task.

4. Under extremely difficult task conditions, subjects will not differ in conformity as a function of self-esteem level.

METHOD

The experiment was carried out in two phases. In the initial phase the focus was on comparing the differential reactions to social influence of female subjects varying in self-esteem under simple and moderately difficult task conditions. All subjects participated under both conditions of task difficulty. However, it was not until the initial phase of the study was almost complete that the decision was made to assess conformity under high-difficulty task conditions. In order to ensure that this phase of the study was comparable to the first with respect to the number of tasks in which subjects participated, these latter subjects were also tested under simple task conditions. In both phases of the study order of tasks was counterbalanced. Thus, the study can be viewed as consisting of two factorial designs, each composed of two levels of task difficulty (low versus moderate or low versus high difficulty) and three levels of self-esteem (low, medium, and high). In addition, certain comparisons can also be made between results obtained in the two phases.

Subjects

Ninety-three females from a small teachers college were administered the Janis and Field (1959) test of self-esteem. From the distribution of self-esteem scores, subjects from the upper, middle, and lower thirds were randomly selected and invited to participate in a study concerning artistic judgment. Of the 70 subjects who subsequently participated in the experiment, 3 were dropped from the data analysis due to their suspicion of the experimental manipulations.

Procedure

Each subject arrived at the experimental site alone, but moments later was joined by a second female, who was introduced as an art major from a neighboring university. The art major was actually an experimental accomplice whose behavior was standard across all experimental conditions. The experimenter introduced the study as stemming from a concern with mediocre standards of artistic excellence in the United States. Relevant to this concern with standards of aesthetic judgment, the two participants were told that they were to make judgments of a series of paintings. During this task the participants were to learn about each other's judgments. This was in preparation for a later confrontation in which their judgments were to be discussed and compared to those made by a group of art experts.

The subject and the accomplice were then led to separate experimental booths. The subject was shown a volume which contained color photographs of 16 pairs of abstract paintings. The paintings were the products of graduate art students in another area of the country and were thus unfamiliar to the subjects. The subject was told that she would have to make comparisons between each pair of paintings on the basis of criteria supplied by the experimenter. She was then shown a panel on which appeared 10 buttons and 10 lights. In each case her judgments were to be indicated by pressing 1 of the 10 buttons. It was also demonstrated to the subject that the panel was mechanically connected to similar panels in possession of the experimenter and the partner, and that by throwing a switch she could communicate her judgments to both of them.

The first task was then described to the subject. It was one of the three following types:

Low-Difficulty task. Some paintings clearly contained more of a color best described as "red-orange" than others. The subject's task was to judge which painting in each pair contained more of the color. The judgment could be indicated by pressing 1 of the 10 buttons. If the painting to the right contained considerably more of the color than the one to the left, the subject was instructed that she might want to press the button on the extreme right of the panel; if the two paintings were relatively similar with respect to coloration, a button toward the center of the panel might be depressed, and so on. It was felt that judgments on this task were highly dependent on physical reality and would thus be accompanied by a high degree of confidence.

Moderate-Difficulty task. For this task the subject was to compare each pair of paintings on the dimension of "creativeness." A button was to be pressed according to which of each pair the subject felt to be the most "original or imaginative." While on one hand persons seem to have fairly well-defined categories for making such decisions (cf. Gergen & Breger, 1965), on the other such judgments are not confidently anchored in physical reality.

High-Difficulty task. Here the subject was to compare each of the paintings on a dimension of "aesthetic goodness." This rating was to be a composite judgment based on four separate subdimensions: proportion or balance among the various forms in the painting, use of color, skillfullness of technique, and the expressiveness of the painting. Since each of these criteria is highly subjective and since making a composite judgment in a rapid period of time presents many additional complexities, it was felt that decisions on this task would be accompanied by a very low level of confidence.

For each pair of paintings in each task in which the subject participated, two judgments were required. She was told that on a signal from the experimenter she and her partner would each turn to the first pair of test paintings. At this signal an electric timer, situated in front of the subject, would be activated. The subject was to make an immediate or "snap" judgment and to press a button to indicate her judgment. When she pressed the button the timer would be stopped. In order to justify this procedure, the subject was told that the study was also concerned with the amount of time required to make immediate decisions about works of art. She was advised to press the button which most closely represented her initial judgment as it would also provide a convenient aid to memory in the discussions following the experiment. However, she was told, the snap decision would be a private one and would register on neither her partner's nor the experimenter's apparatus. In actuality, the experimenter monitored these initial judgments on her panel in the adjoining room. As a distinct advantage over many methods for assessing conformity, the subject's initial judgment could thus serve to establish a base line against which to assess the effects of subsequent social influence pressures.

The subject was then told that the second judgment was to be a "more considered one." It was to be indicated by pressing both the button and a switch at a signal to be given by the experimenter approximately 30 seconds after the snap decision. Throwing the switch at the same time the button was pressed would ostensibly deliver the subject's decision to both the partner and the experimenter. The subject found, however, that the way in which the apparatus was constructed made it necessary for her to see her partner's "decision" in each case before she communicated her judgment to the partner and the experimenter. The "partner's decision" was registered on the series of 10 lights corresponding to the 10 buttons on the subject's panel.

The apparatus was so arranged that the "partner's decision" was, in actuality, under the control of the experimenter. Based on the initial judgment of the subject, the experimenter would systematically determine the amount of discrepancy the subject was to find between her initial rating and the partner's more considered decision. Thus, on 14 of the 15 test pairs the subject found that the partner's "considered decision" differed from her initial decision. In each of these instances the subject found that the partner had pressed a button from one to four positions toward the opposite end of the panel. The magnitude of the discrepancies from trial to trial was preplanned, and this pattern was the same for all subjects and for all task conditions.

After participating in both tasks, the subject answered several questions about the tasks and about possible suspicion of the manipulations. The experiment was then explained in detail, the subject pledged to secrecy, and payment made for participation.

RESULTS

Phase I

As will be recalled, all subjects in this phase of the experiment participated in both the low- and moderate-difficulty tasks. Before examining the major results, it is useful to inquire as to whether the two tasks were perceived to differ in difficulty. All subjects had been asked to rate on a 4-point scale the degree of difficulty they experienced in rating the pictures. Analysis of variance revealed

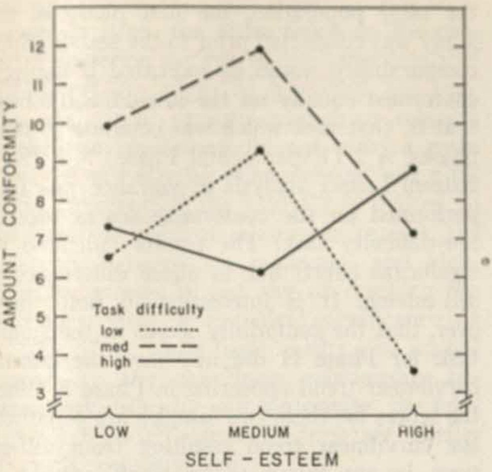


FIG. 1. Conformity as a function of self-esteem level and task difficulty.

that subjects rated the moderately difficult task as significantly ($p < .05$) more difficult than the low-difficulty task. Further confirming evidence was available in the form of the amount of time required by subjects to come to a snap decision regarding the pictures. In the low-difficulty task the mean reaction time (total for all pictures) was 46.8 seconds, whereas in the moderate-difficulty task it was 56.9 seconds; the difference between the two is significant at beyond the .05 level. The manipulation of task difficulty, then, appears to have been quite successful.

There were two major hypotheses to be tested in this phase of the study. It was first predicted that while under simple task conditions self-esteem would not be related to degree of conformity, under moderately difficult task conditions the relationship between self-esteem and conformity would be curvilinear. Further, it was predicted that conformity would be greater when the task was moderately difficult than when it was simple. Conformity was measured by the degree to which the subject would alter her initial judgments about the paintings in the direction of her partner's decisions. For each subject, the discrepancy between the button chosen before and after being informed of her partner's decision was summed for the series of 15 judgments; counterconformity was subtracted from this total.

The results for subjects in this phase appear in Table 1 and are featured graphically in Figure 1. Conformity scores for subjects in the present phase of the study were subjected to a Type I (Lindquist, 1956) analysis of variance. The results of this analysis are found in Table 1. As depicted in the figure, the relationship between self-esteem and conformity under moderately difficult task conditions is in the predicted curvilinear direction. However, contrary to expectation, it is also noted that the same curvilinear relationship is found in the low-difficulty task condition. The significant ($p < .025$) main effect for self-esteem level lends statistical support to the curvilinear relation between self-esteem and conformity. Individual cell comparisons indicate that the greater part of this significance is contributed by the differ-

TABLE 1

MEANS, STANDARD DEVIATIONS,[a] ANALYSIS OF VARIANCE OF CONFORMITY SCORES

Task difficulty	Self-Esteem level		
	Low	Medium	High
Phase I			
Low			
$\bar{X}$	6.54	9.38	3.54
SD	5.95	8.48	6.37
N	13	13	13
Moderate			
$\bar{X}$	9.92	11.92	7.15
SD	6.09	10.08	5.90
Phase II			
Low			
$\bar{X}$	8.64	6.54	6.75
SD	5.25	6.39	5.52
N	11	13	8
High			
$\bar{X}$	7.36	6.15	8.88
SD	6.05	6.07	8.92

Analysis of variance: Phase I

Source	*df*	*MS*	*F*
Between *S*s	(77)		
High, medium, low self-esteem (B)	2	183.58	4.92*
Error (b)	75	37.31	
Within *S*s	(78)		
Easy vs. moderate difficulty (A)	1	197.13	26.28**
A × B	2	.55	
Error (w)	75	7.50	

[a] Hartley test for heterogeneity of variance, *ns*.
* $p < .025$.
** $p < .001$

ence between the medium- and high-self-esteem groups ($p < .001$). The difference between the medium- and low-self-esteem groups is not significant, and, more supportive of the hypothesis, the high- and low-self-esteem subjects do not differ in conformity. In summary, then, the relationship between self-esteem and conformity tends to be curvilinear regardless of whether the task is simple or moderately difficult.

The prediction of greater conformity in the moderately difficult task than in the low-difficulty task was fully supported by the results. In all self-esteem conditions the results were in the predicted direction, and the relevant main effect is significant at beyond the .001 level.

Phase II

In this phase of the study a second group of subjects participated in both the low- and high-difficulty tasks. Independent checks on the efficacy of the manipulations again proved successful. Following the experiment subjects in this phase rated the high-difficulty task as significantly ($p < .05$) more difficult than the low-difficulty task. Completion time for the snap decision also differed appropriately between the two tasks ($p < .05$).

There were again two predictions to be tested. It was first anticipated that there would be no difference in conformity between the low- versus the high-difficulty tasks and, second, that under both low- and high-difficulty tasks there would be no significant relationship between self-esteem and conformity. The means and standard deviations for conformity scores in these various conditions can be found in Table 1. A Type I analysis of variance was again used, and the results of this analysis fully support the expectations. No main effects due to task difficulty or self-esteem nor any significant interactions emerged from the analysis. Indeed, the highest F ratio was only 1.48 ($p < .40$).

Phase I-Phase II Comparisons

Further information can be gained by comparing certain results from the two phases of the study. First it is useful to establish the comparability of the samples used in the two studies. Although subjects were drawn from the same population, the first phase of the study was conducted prior to the second. Such comparability would be indicated if subjects conformed equally on the low-difficulty task, that is, that task which was common to both phases. A 2 (Experimental Phase) × 3 (Self-Esteem Level) analysis of variance was thus performed on the conformity scores for the low-difficulty task. The results indicated no significant effects due to phase differences or self-esteem. It is interesting to note, however, that the conformity scores in the simple task for Phase II did not have the general curvilinear trend appearing in Phase I. When the scores for these two samples are collapsed, the curvilinear trend resulting from self-esteem is not found to be significant ($F < 1.51$). Thus, while scores in the simple task do fall in a curvilinear pattern in Phase I of the study, this pattern fails to replicate in Phase II.

The second comparison of major interest is between conformity on the moderate- versus the high-difficulty tasks. Before examining this difference, however, it is first important to know whether the high-difficulty task was perceived as more difficult than the moderate-difficulty task. Since subjects had rated each task in comparison to the low-difficulty task on only a 4-point scale, a simple comparison of the scale ratings of task difficulty used by the two groups was inappropriate. More appropriate was a comparison of the amount of time spent by each group in reaching the initial or snap decision. A t-test comparison of the mean latency time between the two groups, greater for the high-difficulty than the moderate-difficulty subjects, was just shy of reaching significance ($p < .06$, two-tailed test). In order to shed additional light on the perceived difference between the two tasks, a group of 12 independent raters was exposed to both the high- and moderate-difficulty tasks. Ten of the 12 rated the high-difficulty task more difficult.

In turning to conformity differences between the two conditions, to be consistent with the hypothesis subjects in the moderate-difficulty task should have demonstrated greater conformity than those in the high-difficulty task. To illustrate this difference, the conformity means for the high-difficulty task conditions have been added to Figure 1. As can be seen, the difference is in the predicted direction for two of the three levels of self-esteem. As might be suspected, the two-way analysis of variance comparing subjects in the two conditions did not yield a main effect ($F < 1.60$) for differences in task difficulty. However, if the analysis is limited to the low- and medium-self-esteem groups, the difference between the moderate- and high-difficulty groups is just shy of reaching the customary level of statistical significance ($p < .07$). Thus, while these results give partial support to the curvilinear relationship between task difficulty and conformity, the findings for high-self-esteem subjects are worthy of special consideration.

DISCUSSION

One of the major propositions underlying the present study was that among females the relationship of self-esteem to persuasibility is altered as task difficulty increases. Specifically, it was felt that under very simple and highly complex tasks, differences in self-esteem would not differentially predict conformity; however, under moderately difficult task conditions a curvilinear relationship would emerge between self-esteem and conformity.[3] Indeed, the results in the medium-difficulty task condition do lend support to the curvilinear hypothesis. While the high- and low-self-esteem subjects did not differ from each other in conforming behavior, the medium-self-esteem subjects were most conforming. Also in line with predictions, the curvilinear relationship was not evidenced in the high-difficulty task condition.

The one finding that is at variance with prediction was the curvilinear relationship between self-esteem and conformity in the simple task condition. Although this result did not replicate in the second phase of the experiment, the finding is deserving of attention. Given that the task was perceived as least difficult, and elicited least conformity, the most probable explanation is that the low-difficulty task did not reach simple enough proportions to dampen the effects of self-esteem. Further research might profitably explore the self-esteem–conformity relationship under task conditions of a less complex variety.

It was also predicted that regardless of self-esteem level, the relationship between task difficulty and conformity would be curvilinear. While conformity was expected to increase as the task became more complex up to a certain point, it was expected to abate as the task became extremely difficult. The majority of the data support this prediction. When exposed to the moderately difficult task, subjects conformed significantly more than when confronted with the simple task. In addition, when subjects were exposed to the simple task they were no more persuasible than when working at the highly difficult task. When the low- and medium-self-esteem subjects were considered alone, the comparison of the moderate- versus high-difficulty task was just shy of reaching significance at the .05 level. The one result which is thus at variance with prediction was obtained in the high-self-esteem group. Here greatest conformity was found in the high-difficulty task.

This latter result seems reasonable if one considers more closely the relationship between task difficulty and self-esteem. The psychological counterpart of increasing task difficulty is said to be a decrease in one's feelings of self-confidence. However, it is by no means clear that a standard increment in task difficulty produces an equal decrement in self-confidence across all levels of self-esteem. While the results did not indicate perceptual differences between low- and medium-self-esteem subjects, the finding that high-self-esteem subjects conform most under highly difficult task conditions does suggest a differential perception of the task. Such subjects may be more reticent than others to relinquish their position of enhanced self-confidence. In essence, high-self-esteem subjects may have subjectively felt that the task was only moderately difficult. Evidence suggesting this view was found in the data related to task completion time. In the high-difficulty condition, high-self-esteem subjects more rapidly reached a solution, presumably finding the task easier, than either of the other self-esteem groups in this condition. Unfortunately the statistical analysis reveals that the finding is at best only suggestive.

---

[3] It should be noted that Cohen (1959) has also proposed that both high- and low-self-esteem persons demonstrate certain styles of interpersonal defensiveness. However, the inadequacies of his rationale in the present instance have been discussed by Cox and Bauer (1964).

The present study has implicitly been based on the assumption that subjects were primarily involved in a problem-solving activity. In this sense, either extreme in self-esteem might function to inhibit the seeking out of relevant information from others. On the other hand, in many of those studies that have found a linear relationship between level of self-esteem and persuasibility, the stimulus situations employed may have fostered a greater consideration for social solidarity. The possibility thus emerges that the curvilinear relationship between self-esteem and persuasibility found in the present study may be a function of the particular setting employed rather than the sex of the subjects.[4] A study in which sex, self-esteem, and experimental set are simultaneously considered would seem a next important step in determining the link between self-esteem and social conformity.

---

[4] R. Nisbett and A. Gordon, in a personal communication, November, 1965, report that, under conditions in which the source of influence does not provide supporting arguments for a position taken, a curvilinear relationship is found between self-esteem and persuasibility among both male and female populations.

REFERENCES

ASCH, S. E. Effects of group pressure upon modification and distortion of judgments. In E. E. Maccoby, J. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1958. Pp. 174–182.

BERKOWITZ, L., & LUNDY, R. M. Personality characteristics related to susceptibility to influence by peers or authority figures. *Journal of Personality*, 1957, **25**, 306–316.

BLAKE, R. R., HELSON, H., & MOUTON, J. S. The generality of conforming behavior as a function of factual anchorage, difficulty of task, and amount of social pressure. *Journal of Personality*, 1956, **25**, 296–305.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, G. A. *A study of thinking*. New York: Wiley, 1957.

COHEN, A. R. Some implications of self-esteem for social influence. In C. I. Hovland & I. L. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 102–120.

COLEMAN, J. F., BLAKE, R. R., & MOUTON, J. S. Task difficulty and conformity pressures. *Journal of Abnormal and Social Psychology*, 1958, **57**, 120–122.

COX, D. F., & BAUER, R. A. Self-confidence and persuasibility in women. *Public Opinion Quarterly*, 1964, **28**, 453–466.

GERGEN, K. J., & BREGER, I. Two forms of inference and problems in the assessment of creativity. In, *Proceedings of the 73rd annual convention of the American Psychological Association 1965*. Washington, D. C.: American Psychological Association, 1965. Pp. 215–216.

HOCHBAUM, G. M. The relation between group members' self-confidence and their reactions to group pressures to uniformity. *American Sociological Review*, 1954, **19**, 678–687.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *Journal of Personality*, 1954, **22**, 504–518.

JANIS, I. L., & FIELD, P. B. Sex differences and personality factors related to persuasibility. In C. I. Hovland & I. L Janis (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 55–68.

LESSER, G. S., & ABELSON, R. P. Personality correlates of persuasibility in children. In C. I. Hovland & I. L. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 121–137.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.

LINTON, H., & GRAHAM, E. Personality correlates of persuasibility. In C. I. Hovland & I. L. Janis (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 69–101.

LONDON, P., & LIM, H. Yielding reason to social pressure: Task complexity and expectations in conformity. *Journal of Personality*, 1964, **33**, 75–98.

(Received April 15, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 23-36

THE ANAL CHARACTER:

A REBEL IN THE DISSONANCE FAMILY[1]

FRANCES V. BISHOP[2]

New York University

This investigation maintains that anal characters will not conform to Festinger's dissonance theory, but will base judgments on internal rather than external standards. In a modified replication of Festinger and Carlsmith's (1959) experiment, 67 Ss were preselected for high and low anality. 2 levels of privation were manipulated. Hypotheses predicting an interaction between privation and anality were confirmed. High-anal Ss under conditions of dissonance and high privation responded to boring tasks by increased dislike compared with low anals, and under low privation by increased liking. Predictions for high anals run counter to dissonance theory. An alternate explanation, embracing reactions of both groups, is examined.

In expounding his theory of cognitive dissonance, Festinger (1957) made note of the possible impact of personality differences on the way people react to the existence of dissonance. However, the theory, developed on a nomothetic model, is not centrally concerned with individual differences and does not pretend to deal with them systematically. This paper will attempt to focus on such differences. It will examine the mediation of dissonance in connection with the personality constellation of the anal character.

Festinger (1957) holds that the magnitude of dissonance varies inversely with the degree of justification for performing incongruent behavior; that is, if one has little reason to perform distasteful behavior, the discomfort in doing so is relatively great. The greater the discomfort, the greater the drive to reduce it. Behavior, once performed, cannot be undone; therefore, to attain consonance, the element most likely to undergo change, dissonance theorists claim, is one's attitude. This leads to the paradoxical prediction that the more inconsequential the inducement to perform discrepant behavior, such as an aversive task, the more a person will like that task.

An experiment by Festinger and Carlsmith (1959) put this prediction to the test. All subjects were paced through two tasks which a control group judged as boring. Subjects in the high-dissonance condition were offered a low reward ($1) for telling the incoming subject (a confederate) that the experiment was "fun." In the low-dissonance condition, subjects were offered $20. The investigators predicted and found that the smaller the reward, the more subjects liked the tasks; that is, where justification for talking to the stooge was low, dissonance occurred, which was reduced by a positive change in opinion about the boring tasks.

Since arousal of dissonance is an intraindividual phenomenon, it seems appropriate to attempt to clarify such paradoxical results by recourse to knowledge about personality. A reasonable assumption is that interaction between experimenter and subject (Rosenberg, 1965) might account at least in part for such unexpected outcomes. The forced-compliance paradigm suggests commitment by the subject to disagreeable behavior under conditions of indirect pressure from an authority figure. The very notion of compliance presupposes interpersonal interaction and furnishes forced-compliance experiments with a toehold in the field of personality. Thus, certain subjects might be expected to yield to the indirect demands of an experimenter, while others would resist such demands. Dissonance theory has nothing to say about who will be a "yielder," who a "nonyielder." Freud (1959a), however, in his description of the anal character has provided us with the prototype of the nonyielder or oppositional personality.

---

[1] Based on a doctoral dissertation prepared for the Department of Psychology, Graduate School of Arts and Science, New York University. The author wishes to thank Irving Sarnoff, sponsor, Philip Zimbardo, and Leo Goldberger for their encouragement and assistance.

[2] Now at the Olive View Hospital, Olive View, California.

Freud was the first (1938, 1959a) to infer a systematic relationship between a triad of adult characteristics—obstinacy, parsimony, and orderliness—and early childhood training during the postulated psychosexual stage of anal development. That these three traits are correlationally connected has been shown by Finney (1963) in a study of the anal character in children. Freud (1959a) linked obstinacy to interest in defecation on the grounds that infants could behave with "great self-will about parting with their stools [p. 46]." He held that interest in money is a sublimation of, and orderliness a reaction formation against, interest in feces.

Since Freud's early description, the constellation of traits attributable to the anal character has been expanded. Certain writers, such as Jones (1961), have attempted to show that orderliness, by extension, may result in clarity of thinking. This notion is in accord with the finding that the obsessive-compulsive character, said to develop from the anal stage, usually obtains high *F*+ scores on the Rorschach (Rickers-Ovsiankina, 1960; Rorschach, 1942), indicating not only good reality contact but also the capacity to structure ambiguous material with competence and clarity.

Fenichel (1945) maintained that a genetic relation exists between time and feces as well as between money and feces. He claimed that the child acquires the ideas of order and disorder regarding time from the intervals at which defecation takes place and from how long the process takes. Fenichel held that disturbances in anal personality may lead to stinginess or prodigality with either time or money and sometimes to both attitudes alternately.

Another tendency of the person high in anality is to internalize the values expressed in social training as he takes pleasure in mastering anal impulses. Thus, the anal-retentive (as opposed to the anal-expulsive) type is far more likely to lean on his own value system than to depend on outer standards. Couch and Keniston (1960) postulated and found marked personality differences between yeasayers and naysayers. After completing a clinical study on the two types, they attributed the differences to opposite resolutions of the anal problems around control of impulses. In their opinion, the negativism of the anal-retentive character (the naysayer) is closely linked with a tendency to repudiate external standards and to internalize controls. Therefore, one should expect less compliance and greater adherence to his own views from this type of person than from others.

The above description of the anal character is not intended to be exhaustive, but lists primary anal traits relevant to the present study. Freud (1920) has specified the conditions that might render those traits most pronounced. In 1915, after postulating specific stages of infantile development, Freud predicted that intense gratification or frustration at any single stage would lead to certain enduring character patterns, and that, given anxiety or stress, an individual would tend to regress to the stage at which gratification or frustration was greatest. Under conditions where supplies are taken from the anal character (conditions of privation), a mild regression to the anally fixated stage should occur, anal traits should be heightened, and such a person should respond negatively to privation. Under conditions of reward, the potential reactions of the anal type are less clear. Nevertheless, where tension is created by the possibility of winning money, one can speculate that the anal character should find gratification in reward through alleviation of tension, in accordance with Freud's constancy principle.

PINPOINTING THE SILENT VARIABLES

To demonstrate how the anal traits just described might have a bearing on the mediation of dissonance, it is necessary to point up certain variables that operate silently in forced-compliance experiments, yet interact with personality variables. These "silent" variables are:

1. *An outer reference point,* identified by the position of the experimenter, usually at odds with that of the subject. In some experiments, this reference point is recognized and made operational by a bogus consensus or falsified opinion (Asch, 1951; Deutsch &

Gerard, 1955; Fisher & Lubin, 1958; Zimbardo, 1960). In many other experiments, however, it is implied but not specified.

Such an implied reference point exists in the Festinger and Carlsmith (1959) experiment cited above. After the subjects had conversed with the confederate, but before they rated attitudes toward the tasks, the experimenter said, "I hope you did enjoy it [the experiment]. Most of our subjects tell us afterward that they found it quite interesting [p. 206]." Although this remark, as far as the students were concerned, was quite casual, the investigators point out that it was made to facilitate a change in attitude. It is the type of hidden reference point at issue here. The person high in anality, accustomed to rely on his own value systems, should be less influenced than others by such unidentified or "silent" reference points.

2. *Pressure to yield to the experimenter* is usually exerted by the experimenter's comments, instructions, and by his requests to the subject to perform discrepant behavior. The subject usually is faced with the choice of pleasing the experimenter and getting on with the experiment or of refusing to perform the discrepant behavior, thus sabotaging the research. Usually the pressure is great, the choice minimal.

3. *Ambiguity* occurs at several points in the forced-compliance paradigm, particularly in connection with the two variables just enumerated. Certainly, ambiguity is at its height where the subject is harnessed into a behavior he dislikes while led to believe he has free choice. Ambiguity also occurs at the point of final attitude rating where the subject is confronted by discrepant cognitions between what he once thought and just now performed. Here, the postbehavioral ambiguity stems from the fact that consistency can be derived in two ways—by adhering to one's original stand, or by changing one's opinion to conform to one's behavior. Since dissonance theory postulates a drive to reduce inconsistency, in such situations the subject is driven in two directions. This opens a wedge for individual differences to emerge. Dissonance theory deals with only one mode of gaining consonance—by aligning opinion with behavior—on the grounds that behavior, once performed, cannot be undone, and therefore resists change more than attitude. The theory bypasses the second mode of gaining consonance—that of holding to one's initial position.

From the foregoing analysis, it would seem that the anal character, guided primarily by internal value systems, should reject external standards; considered oppositional, should pit himself against pressure; described as orderly in thought, should meet ambiguity with clarity; being parsimonious, should respond negatively to privation; and being obstinate, should adhere to his own opinion.

Thus, it would appear that there are pockets in the dissonance experimental paradigm in which subjects high in anal character traits would react in ways that differ from other subjects. Therefore, if one were to create experimentally a situation in which persons high in anality were led to expect a high reward for performing a disagreeable task and were then subjected to relatively drastic privation, one would expect, from psychoanalytic theory, that they would respond to privation with marked dislike and would express a more disgruntled opinion than those for whom reward and privation had less intense meaning.

Such a prediction, when applied to the forced-compliance paradigm, would be opposite to predictions from dissonance theory. For dissonance experimenters (Brehm & Cohen, 1962; Festinger & Carlsmith, 1959) have shown that tension created by dissonance leads subjects in low-reward conditions to like an aversive task better than do those in high-reward conditions. This poses the problem of hierarchy of motivation in selecting the tension to be reduced. It is held here that anal characters, under conditions of mild anxiety or privation on the one hand, or gratification on the other, will respond primarily to instinctual needs, in accordance with Freudian theory, and that their cognitive needs will differ from those postulated by dissonance theory.

HYPOTHESES

The following hypotheses derive from the above review:

Hypothesis I

Under conditions of forced compliance and high privation, persons high in anality will tend to dislike a disagreeable task more than persons low in anality.

Hypothesis II

Under conditions of forced compliance and low privation, persons high in anality will tend to like a disagreeable task more than persons low in anality.

It should be noted that these predictions are interrelated, bear on the expectation of an interaction between privation and anality; and run counter to dissonance theory for those high in anality.

No enduring pattern of characteristics has been recognized for those low in anality. Consequently, no a priori predictions from personality theory can be made for such subjects. The best available prediction is that subjects low in anality should conform to dissonance theory.

Hypothesis III

From psychoanalytic theory a prediction also can be made that subjects high in anality will show more reluctance about giving up time than those low in anal traits. In other words, under conditions of both high and low privation, persons high in anality will be less willing to donate time to an experiment than those low in anality.

METHOD

The present experiment, based on a 2 × 2 design with two control groups, was modeled on the investigation by Festinger and Carlsmith (1959) described earlier. However, the current version was restructured to accommodate two new variables, privation and anality, and two control groups, one for high and one for low anality. The Festinger and Carlsmith (1959) study had been criticized by Janis and Gilmore (1965) on the grounds that the high reward was implausible and engendered suspicion. To counteract this, a chance system of payment (rigged) was devised, in the form of a lottery, to be described, along with other modifications, in the next section.

A major methodological concern centered around devising a tool that would heighten and educe the full panoply of the anal character in the laboratory. While it was expected that many anal traits would be brought into play in connection with dissonance, only one, parsimony, was experimentally elicited in the laboratory. This was done by creating two conditions of privation, measured operationally by establishing differences between reward expectation and actual rewards received. These manipulations are described in the section on procedures.

Subjects were premeasured for anality [3] by selecting high and low scorers on the Ah or hoarding behavior scale of Grygier's (1956) Dynamic Personality Inventory. This inventory consists of 325 items, subdivided into scales associated with specific patterns of psychosexual development, such as oral, anal, and phallic patterns, as well as with other personality features. Subjects are asked to indicate whether they like or dislike each item in the inventory.

Although there are six anal scales within the inventory, only the hoarding behavior scale was used.[4] High scores on this scale measure "hoarding behavior, anxious possessiveness, and stubborn, clinging persistence." [5] Low scores indicate "lack of concern with material possessions, extravagance" (see Footnote 5).

The hoarding behavior scale is composed of the following items, randomly distributed throughout the test: Item 26: undoing all knots (rather than cutting the string); Item 89; collecting stamps; Item 106: hanging on to old clothes you never wear; Item 124: spending money on passing pleasures to have fun; Item 126: keeping your receipts long after you have paid the bills; Item 163: storing many old things in case you need them one day; Item 209: saving boxes or medicine bottles; Item 212: collecting books, journals, or magazines; Item 241: collecting coins; Item 288: collecting postcards from your holiday; Item 296: keeping a large stock of canned goods in your cupboard.

Percentile cutoff points for admission to the sample were held at stringent levels to compensate for the fact that no other effort to match personality variables was envisioned. The criterion for high anality was set at the 77.3 percentile of Grygier's standardization group for the Ah scale and at the 18.66 percentile for low anality.

---

[3] Anality, as used in this investigation, refers to the anal-retentive as opposed to the anal-expulsive type.

[4] The Ah scale was selected for the follow[illegible]easons: (*a*) According to Freudian theory, hoarding behavior and stubborn persistence are typical anal characteristics and are the most forthright derivatives of interest in feces; (*b*) the items on this scale are unambiguous and have relatively clear operational definitions; (*c*) an original attempt to use all six anal scales had to be abandoned because extensive pretesting yielded an extremely small number of subjects, the reason being, according to a personal communication from Grygier, that the Ah scale tends to correlate negatively with all other anal scales with the possible exception of the Aa scale (submissiveness to and rejection of authority) and, to some extent, the Ai scale (insularity, reserve, and mistrust).

[5] Quoted from Grygier's unpublished manual, distributed for research purposes. A copy of this manual may be found in an appendix to the doctoral dissertation on which this article is based (cf. Footnote 1).

The final sample consisted of 67 male undergraduates from New York University, with 35 high and 32 low anals almost evenly distributed in the different conditions.[6]

Nineteen protocols were excluded from the final sample. Prior to eliminating data, the experimenter and another doctoral candidate rated all protocols for disqualifying factors. In 11 cases, the decision to discard data was based on scientifically mandatory reasons, such as a subject's refusal to make the discrepant statements, participation in concurrent deception research, regarding the lottery as deliberately rigged, and similar factors. In 8 other less-clear-cut cases, the decision was made to drop protocols because subjects were suspicious of the lottery. An analysis of variance was computed that included not only protocols of all subjects retained in the final sample but also included protocols of the 8 subjects who suspected the lottery. This analysis yielded results in the same direction and at virtually the same levels of significance as for the final sample. That is, for the enlarged spurious sample, the interaction between anality and privation for the major hypotheses reached a significance level of $.10 > p > .05$. For all other variables, significance levels were identical in both samples.

Procedure

Two boring tasks from the original experiment were retained, each lasting 20 minutes. One task consisted of turning each peg in a 48-peg board one-quarter turn clockwise, then repeating this process; the second consisted of mounting 100 spools on spindles, emptying them into a tray and continuing this procedure until told to stop. The experimenter clocked the subjects for 45 seconds, 3 minutes after starting and before ending each task.

Whenever feasible, instructions were played from a [illegible] recorder. Subjects were told the experiment was sponsored by the National Research Foundation and that its purpose was to analyze industrial performance tasks. The experimenter added that, because of budget restrictions, funds for paying volunteers had been pooled and a lottery established, giving subjects the opportunity to win a larger payment than if the money were evenly distributed. The experimenter indicated that the subject might also be a loser.

Before participating in the lottery, the subject was shown two dummy lots (one entitling him to receive nothing, the other entitling him to $1). While he watched, these were sealed with double stick-tape and placed among the other lots so they would adhere to the box. Each subject drew a lot that led him to expect potential maximum winnings of $20 and gave him a chance to participate in the final draw. He received 20 $1 coupons and held these until the tasks were completed. The fact that the coupons were redeemable for money was emphasized. When the tasks were over, the subject was given a punchboard for the final draw. This was rigged in such a way that, although a group of random numbers were visible representing (false) punches made by other subjects, he could only punch either $19 or $1, depending on experimental condition. Thus, he was forced to relinquish either $1 (low privation) or $19 (high privation).

The experimenter then gave the subject a bogus explanation of the experiment, telling him it was designed to test the influence of attitude on performance. One group (Group A), the experimenter added, was given no information about the tasks, but the second group (Group B) was led to believe the tasks were "fun." The experimenter then observed that such statements would be more convincing coming from the student than coming from the experimenter. She handed the subject a sheet of paper entitled, "Instructions for Group B," on which were typed a number of complimentary remarks about the experiment. The subject was asked to phrase these remarks in his own language and to record them as spontaneously as possible on a tape recorder. He was informed that his statements would be played back to the "incoming subject."

The guide sheet contained the following statements:

> The experiment was very enjoyable, I had a lot of fun; I enjoyed myself, it was very interesting, it was challenging, it was intriguing, it was exciting.

Except for addition of the phrase, "it was challenging," the statements duplicated the Festinger and Carlsmith guide sheet. However, *tape-recording* this message was a departure from the original model, in which each subject confronted the incoming student (a stooge) in person. Another divergence consisted in the fact that in the original version subjects were specifically paid to deliver the discrepant message, whereas in the present version the reason for payment was not specified other than for participation in the experiment as a whole.

Usually subjects readily complied with the experimenter's request to make the misleading statements. Where there was hesitation, compliance was induced in three standard probes. Just before the subject recorded the discrepant statements, the experimenter again reminded him he was recording of his own free will.

When the recording was completed, the experimenter thanked the subject for his participation and informed him that most subjects, when the experiment was over, said they liked it very much. (This statement and the experimenter's request to the subject are identified here as the outer reference points.)

---

[6] The original plan called for a larger sample, but the experiment was brought to an abrupt halt due to theft of two tape recorders, taped data, taped instructions, and other equipment used in connection with subjects.

The dependent variables were then presented to the subject as a survey conducted by the University and not associated with the experiment proper. A complete disassociation of this survey from the experiment, as performed by Festinger and Carlsmith (1959), was not feasible due to manpower and space shortages. However, the experimenter attempted to stress such a disassociation by describing herself as an "outsider." She assured the subject that the survey would be anonymous, left the office while it was being filled out, and suggested to the subject that he seal it and take it himself to the department concerned.

Control groups were subjected to quasi-identical treatment, except that they recorded only facts on tape (name, age, class, occupation, number of coupons won) and did not participate in the second draw until the experiment was over.[7] Both experimental and control groups answered a new and extensive postexperimental questionnaire, signed a pledge to silence for which they were paid $1, and were then fully dehoaxed.

ASSESSING THE EXPERIMENTAL MANIPULATIONS

Independent checks to assess the experimental manipulations probed for boredom of tasks, credibility of lottery, privation "take," feelings of choice about making discrepant statements, plausibility of rewards, arousal of dissonance, and other factors. These checks tended to show that the experimental inductions were successful.

Boredom of Tasks

Two control groups, a high-anal and a low-anal control group, rated the tasks as boring or enjoyable on a scale ranging from 0 (extremely boring) to 50 (extremely enjoyable), with 25 at the neutral or midpoint. The mean for the high-anal control group ($M = 23.55$) hovered near the midpoint, but was slightly negative; the label closest to this mean was "mildly boring." This is in line with the ratings by the control group in the Festinger and Carlsmith (1959) experiment, where the mean was "only slightly on the negative side" of the neutral point.

For the low-anal group, on the other hand, the average rating was 14.40, and the closest label was "very boring." Because of the 9-point spread between average scores of both control groups, a t test between means was performed, but fell short of significance ($t = 2.069, .10 > p > .05$). Thus, while both control groups may be considered to have appraised the tasks negatively, it is legitimate to speculate whether dissonance effects might have been more readily perceptible had the high-anal subjects been subjected to tasks they judged as more boring. This issue will be taken up in the discussion section. Table 4 lists average ratings for all groups.

Privation "Take"

Three questions on the postexperimental questionnaire probed for feelings of privation, and for all three the main effect of privation was highly significant. For reasons of psychological closure, control groups were paced through the privation-reward manipulations *at the end of the experiment* and were subdivided accordingly for analysis of variance of all postexperimental questions. This is reflected in the tables below. Since results on all three questions tend to parallel each other and are discussed in detail elsewhere (see Footnote 7), of this batch only the following question will be analyzed here: "In relation to giving up coupons, did you feel or not feel disappointed? To what degree?" The question was rated on an 8-point scale ranging from 0 to 7 with 0 at the extreme positive and 7 at the extreme negative end of the scale.

Table 1 indicates that the main effect of privation was statistically significant beyond the .001 level of significance. It can be seen from the table that means of all groups in the high-privation condition were higher than corresponding means in the low-privation condition and that feelings of disappointment were differentially induced at statistically sig-

TABLE 1

ANALYSIS OF VARIANCE AND GROUP MEAN SCORES DENOTING DEGREE OF DISAPPOINTMENT IN RELINQUISHING COUPONS

Source	*SS*	*MS*	*F*
Anality (A)	18.18		6.65*
Privation (B)	79.46		29.06***
Discrepancy (C)[a]	22.32		8.16**
A × B	.21	.21	.08
A × C	8.70	8.70	3.18
B × C	4.27	4.27	1.56
A × B × C	.94	.94	.34
Within	161.32	2.73	
Total	286.87	4.35	

Condition	*N*	*M*	*SD*
High privation			
High anals			
Experimentals	13	5.08	1.14
Controls	5	3.80	2.14
Low anals			
Experimentals	12	3.33	1.80
Controls	4	3.75	1.64
Low privation			
High anals			
Experimentals	11	3.09	1.50
Controls	6	.83	.69
Low anals			
Experimentals	10	1.80	1.94
Controls	6	1.17	1.07

Note.—Scale ranged from 0 to 7, with 0 at the extreme positive and 7 at the extreme negative end of the scale. (The higher the score, the greater the disappointment.)

[a] (Exp. vs. Con.)

* $p < .05$, $df = 1, 59$.

** $p < .01$, $df = 1, 59$.

*** $p < .001$, $df = 1, 59$.

---

[7] Complete procedures, instructions, and other details are available in the thesis submitted to New York University.

nificant levels between the two conditions regardless of personality type, even though these feelings were more pronounced for the high- than for the low-anal groups. The most disappointment occurred in the high-anal–high-privation experimental group ($M = 5.08$); the least disappointment among experimental groups occurred in the low-anal–low-privation condition ($M = 1.80$). A Duncan multiple-range test of mean differences with unequal N (Kramer, 1956) indicates that the null hypothesis can be rejected at the .001 level.

The main effects of anality and discrepancy[8], likewise were significant beyond the .05 and beyond the .01 levels of significance, respectively. That is, there were significant differences not only among those in the high-privation condition as opposed to the low, but also among high anals as opposed to low anals, and among control groups as opposed to experimental groups.

Credibility of Lottery

It will be recalled that protocols were analyzed and rated for degree of suspicion and that protocols of all subjects who believed without equivocation that the lottery was rigged were discarded. Protocols of highly suspicious subjects also were excluded. The question that bears most directly on the issue of credibility was: "How did you feel about drawing lots as compensation for your time? Did you feel it was fair or not fair?" No statistically significant differences occurred between groups. All groups judged the lottery fair; no mean scores fell in the unfair range.

A second question related to credibility was: "In view of the fact that payments were made within the context of chance drawing of lots, did you feel your compensation was appropriate or inappropriate? To what degree?" All groups in the high-payment condition classified their winnings between "very appropriate" and "moderately appropriate." Thus, all groups perceived this manipulation as intended without regard to differences in personality or in experimental treatment.

Still another question sought to ferret out attitudes to the lottery: "In the future would you favor or not favor the pooling of funds and drawing of lots to pay for research subjects?" No statistically significant differences occurred between groups. Preponderantly subjects favored the pooling system. On an individual basis, more than 68% gave a positive response.

TABLE 2

ANALYSIS OF VARIANCE AND GROUP MEAN SCORES DENOTING FEELINGS OF CHOICE ABOUT MAKING DISCREPANT STATEMENTS

Source	*SS*	*MS*	*F*
Anality (A)	4.06	4.06	1.26
Privation (B)	4.57	4.57	1.42
Discrepancy (C)[a]	7.88	7.88	2.45
A × B	2.21	2.21	.69
A × C	19.69	19.69	6.13*
B × C	.39	.39	.12
A × B × C	12.46	12.46	3.88
Within	189.49	3.21	
Total	240.75	3.65	
Condition	***N***	***M***	***SD***
High privation			
High anals			
Experimentals	13	1.08	1.44
Controls	5	.20	.40
Low anals			
Experimentals	12	.83	.99
Controls	4	4.25	.83
Low privation			
High anals			
Experimentals	11	1.73	2.05
Controls	6	1.67	2.13
Low anals			
Experimentals	10	1.50	1.91
Controls	6	2.33	2.49

Note.—Scale ranged from 0 to 7, with 0 at the extreme positive and 7 at the extreme negative end of the scale. (The lower the score, the greater the feeling of free choice.)

[a] (Exp. vs. Con.)

* $p < .05$, $df = 1, 59$.

Choice about Making Discrepant Statements

A variety of questions canvassed for feelings of choice and perception of pressure connected with participating in the experiment as a whole or in its various aspects.

The most relevant issue is whether or not the experimental groups felt they were exerting their own free will when they commited themselves to make discrepant statements. Evidence for this issue will be presented here. Additional questions regarding choice and pressure are discussed elsewhere.[9] The following question was asked: "How much choice did you feel you had about talking into the tape recorder?"

Table 2 reveals that mean scores of all groups except one control group cluster between "utter free choice" and "moderate free choice." (It will be recalled that control groups were offered less choice than experimental groups for this manipulation.)

A tally of responses on the positive side of the scale indicates that more than 83% of the subjects perceived themselves as having free choice.

A glance at Table 2 shows there are no significant main effects. Thus, it can be concluded that all experimental groups experienced free choice about making the misleading statements.

---

[8] The variable of discrepancy taps the main effect of making a taped recording that misrepresented the boring quality of the tasks to other potential subjects. Such misleading statements were recorded on tape by all four experimental groups, but *not* by control groups. The latter recorded only facts on tape, such as name, age, class, etc. Because of this difference in manipulation between experimental groups and controls, all tables relating to the post-experimental questionnaire contain the notation, (Exp. vs. Con.), an abbreviation for experimentals versus controls.

[9] Details in thesis.

Dissonance "Take"

To assess differential arousal of dissonance, subjects were asked, "Did you feel honest or dishonest while recording on tape? To what extent?"

Results, reported in Table 3, showed a significant main effect for discrepancy ($F = 47.204$, $p < .001$). That is, control groups, asserting only facts on tape, expressed far greater feelings of honesty than experimental groups involved in making misleading statements. No other main effects or interactions were significant.

For experimental groups, scores cluster between the labels "slightly honest" and "moderately dishonest" but there were no statistically significant differences between means.

RESULTS

Test of the Two Major Hypotheses

The following variable tested the major hypotheses: "How did you feel about the tasks involving the spools and pegs? To what extent did you feel the spool and pegs tasks were boring or enjoyable?" This question was rated on a 51-point scale, with 0 at the negative and 50 at the positive extreme, labeled at 10-point intervals from extremely boring to extremely enjoyable.

TABLE 3

ANALYSIS OF VARIANCE AND GROUP MEAN SCORES DENOTING FEELINGS OF HONESTY WHILE RECORDING ON TAPE

Source	*SS*	*MS*	*F*
Anality (A)	.39	.39	.12
Privation (B)	12.00	12.00	3.62
Discrepancy (C)[a]	156.34	156.34	47.20***
A × B	10.87	10.87	3.28
A × C	.03	.03	.01
B × C	6.12	6.12	1.85
A × B × C	6.44	6.44	1.95
Within	195.44	3.31	
Total	375.40	5.69	

Condition	*N*	*M*	*SD*
High privation			
High anals			
Experimentals	13	4.62	2.13
Controls	5	.60	.49
Low anals			
Experimentals	12	3.33	1.70
Controls	4	.25	.43
Low privation			
High anals			
Experimentals	11	2.73	2.18
Controls	6	.33	.47
Low anals			
Experimentals	10	3.90	2.12
Controls	6	.33	.47

Note.—Scale ranged from 0 to 7, with 0 at the extreme positive and 7 at the extreme negative end of scale. (The lower the score, the more honest the subject felt.)

[a] (Exp. vs. Con.)

*** $p < .001$, $df = 1, 59$.

TABLE 4

ANALYSIS OF VARIANCE AND GROUP MEAN SCORES INDICATING BOREDOM OF TASKS

Source	*SS*	*MS*	*F*
Anality (A)	207.60	207.60	1.76
Privation (B)	441.11	220.55	1.87
A × B	890.58	445.29	3.78*
Within	7193.44	117.93	
Total	8808.45	133.46	

Condition	*N*	*M*	*SD*
High privation			
High anals	13	9.69	11.76
Low anals	12	16.50	9.28
Low privation			
High anals	11	22.18	11.57
Low anals	10	13.90	9.74
Control			
High anals	11	23.55	7.94
Low anals	10	14.40	11.18

Note.—Scale ranged from 0 to 50. (The lower the score, the more boring the task.)

* $p < .05$, $df = 2, 61$.

To detect the interaction between privation and anality, analyses of variance were computed, using the harmonic-mean method for unequal N (Snedecor, 1956). Table 4 indicates that the predicted interaction attained significance beyond the .05 level of significance. That is, under high privation, the high anals obtained a mean score of 9.69, considerably below the average of 16.50 scored by the low anals in the same condition. Under low privation, however, the reverse was found, the high anals reporting the tasks less boring ($M = 22.18$) than the low anals ($M = 13.90$). While neither of these differences was reliable, the mean rating of the high-anal–high-privation group ($M = 9.69$) differed significantly ($p < .01$) from that of the high-anal–low-privation group ($M = 22.18$) as well as from that of the high-anal control group ($M = 23.55$, $p < .01$). Thus, subjects differing in anality responded both differentially and in opposite directions to the quantitative variations in privation. Although the responses among the low-anal groups fell in the direction predicted by dissonance theory, Duncan's multiple-range test extended by Kramer (1956) and applied to means of the three low-anal groups only failed to disclose statistically significant differences. None of the main effects was significant.

It should be noted that mean scores of all groups, including controls, fell below the midpoint (judged the tasks boring), and variations were relative within this range.

Third and Exploratory Hypothesis

The dependent variable testing the third hypothesis was: "How did you feel about having donated time to this experiment? Was it worthwhile or not worthwhile?" This question likewise was measured on a 51-point scale and results are shown in Table 5.

The means for the high anals in both privation conditions ($M = 25.54$ for high privation and $M = 36.45$ for low privation) were both lower than the means for the low-anal groups in the same conditions ($M = 32.17$, 38.30 for high and low privation, respectively), indicating that the high anals were more reluctant about donating time than the low anals. Thus, a trend in the predicted direction was present. However, it was not significant.

TABLE 5

ANALYSIS OF VARIANCE AND GROUP MEAN SCORES ON WHETHER DONATING TIME TO EXPERIMENT WAS WORTHWHILE

Source	*SS*	*MS*	*F*
Anality (A)	18.33	18.33	.17
Privation (B)	822.44	411.22	3.91*
A × B	399.27	199.63	1.90
Within	6417.04	105.20	
Total	7746.48	117.37	
Condition	*N*	*M*	*SD*
High privation			
High anals	13	25.54	13.98
Low anals	12	32.17	8.93
Low privation			
High anals	11	36.45	7.22
Low anals	10	38.30	8.37
Control			
High anals	11	36.91	6.84
Low anals	10	31.60	10.64

Note.—Scale ranged from 0 to 50. (The lower the score, the more negative the reaction.)
* $p < .05$, $df = 2, 61$.

The main effect of privation was statistically significant at the .05 level of significance; that is, for both low and high anals, the higher the privation, the more both groups felt that donating time to the experiment was not worthwhile.

Further Results Pertinent to the Predicted Interaction

Eight more questions were included in the dependent-variable survey. These were formulated in part to explore negative or positive reactions of subjects over a wider range, in part to distinguish between the boring tasks and other aspects of the experiment, thus heightening the subjects' sensitivity to the crucial measures. These questions were:

1. How did you feel about what you said on tape? To what extent did it represent your views? [This variable attempted to probe for any differences between the subject's private view and his public statements.]
2. How would you describe your global reaction to this entire experiment? To what degree was it enjoyable or boring? [This question was included since the misleading statements on tape referred to the experiment as a whole.]
3. How did you feel about your experience of drawing lots? Did you feel this was or was not an exciting experience? [It was expected that subjects would find the lottery more exciting than the boring tasks, and an effort was made to distinguish between these two aspects of the experiment.]
4. How did you feel about your experience of drawing lots? To what extent did you consider it an interesting or uninteresting experience?
5. How did you feel about the tasks involving the spools and pegs? To what extent did it[10] engage or not engage your interest? [Pilot studies had shown that subjects made a distinction between boredom of and interest in the tasks.]
6. How did you feel about your experience of *recording* on tape? To what extent did you feel it was or was not boring or enjoyable?
7. Would you be willing or unwilling to record again under the same or similar circumstances?
8. Would you be willing or not to recommend such an experiment to another volunteer?

For three of these questions, the interaction between privation and anality was significant at the .05 or .01 level, and across all these variables the high-anal group responded more negatively in the high- than in the low-privation condition and more unfavorably

---

[10] A typographical error that appeared on the survey as presented to subjects. "It" should be "they."

TABLE 6

RESULTS OF ANALYSIS OF VARIANCE REGARDING SUBJECTS' ATTITUDES TOWARD VARIOUS ASPECTS OF EXPERIMENT AS SHOWN IN EIGHT ADDITIONAL DEPENDENT VARIABLES

Dependent variables	*F* scores for main effects and interaction		
	Anality	Privation	Interaction (A × P)
1. Discrepancy between public statement and private view	.04	1.34	8.45**
2. Global reaction to experiment	.003	5.98**	4.13*
3. Excitement of lottery	1.86	3.58*	5.01**
4. Interest in lottery	.73	4.92**	1.12
5. Interest in tasks	.17	4.16*	1.15
6. Boredom of *recording* on tape	.20	3.83*	.70
7. Willingness to record again	.01	1.75	1.15
8. Willingness to recommend experiment	.002	.07	.51

Note.—For all variables, regardless of significance, the high-anal groups responded in the direction predicted. Where the main effect of privation was significant, both high- and low-anal subjects responded negatively to high privation, except in reacting to Variable 3, excitement of lottery. For Variables 1, 3, and 8, the direction of response for privation followed the bipolar predictions of the major hypotheses for both personality groups.

* $p < .05$.

** $p < .01$.

than the low anals in the high-privation condition (see Table 6).

DISCUSSION

The significant interaction between privation and anality brings cogent evidence to bear on the validity of the major hypotheses. The reactions of the high- and low-anal groups were bipolar, as predicted. Responses of the high-anal group to high reward were uniformly positive or unchanged in relation to the corresponding control group, whereas their reaction to high privation was uniformly negative, occurring over the range of all 10 dependent variables, in accordance with Freudian rather than with dissonance theory. While no group conformed reliably to predictions from dissonance theory, a tendency in that direction was detected among the low anals, but was not statistically significant.

Before the final appraisal of these results, the following issues deserve examination: (*a*) What weaknesses stem from deviations from the original experiment? (*b*) Was dissonance aroused? (*c*) Were alternate avenues of dissonance reduction selected? (*d*) Do other explanations fit the data with equal or greater cogency?

One of the major deviations from the original experiment consisted in rewarding subjects in the current study by lottery. Thus, after the tasks were completed, but before the misleading statements were made, the subjects engaged in a punchboard manipulation and received coupons that entitled them to either high or low pay to be collected at the end of the experiment. A legitimate criticism, pointing up a weakness in the current version in contrast to the original, is that no explicit connection was made between the promised payment and the making of the discrepant statements. It is possible, therefore, that the subjects might have considered themselves rewarded for performing the boring tasks rather than for making the misleading statements.

Such a situation would not be considered a good test of the dissonance hypothesis, although the Freudian hypothesis might still be operative, because the differential evaluation of the tasks would be greatest immediately after the subjects learned they would receive high or low pay. This would leave the low-pay group more positive toward the tasks than the high-pay group. Therefore, making subsequent discrepant statements would arouse more dissonance in the high-pay group, and they would then presumably have a greater tendency to change their evaluation positively than those in the low-pay group.

Now, if the above criticism were carried to its logical conclusion, then those receiving low pay would evaluate the tasks more positively, experience little or no dissonance in making the discrepant statements, and end up with a positive attitude; those receiving high pay would experience little dissonance after performing the tasks, and would continue to consider them boring. However, they would then experience dissonance in making the misleading statements and would reduce dissonance by finding the tasks more enjoyable. Thus, all groups, regardless of condition, should wind up with a more positive attitude about the tasks than the control groups, an outcome that clearly did not occur. Even if one allows for operation of the Freudian principle among the high anals in the high-privation group, the alternate dissonance explanation cannot account for the significant interaction of the major dependent variable and the bipolar reactions of the different personality groups in the different conditions.

Moreover, further recourse to the data does not support the above alternate explanation. The evidence suggests that subjects, in responding to the dependent variables, linked privation to the entire experiment, not to the boring tasks alone. Two variables bear on this issue. In the first, subjects were asked: "How would you describe your reaction to this entire experiment? To what degree was it enjoyable or boring?" Results in Table 6 show not only that the interaction between privation and anality was significant, but also that the main effect of privation also was significant beyond the .01 level, indicating that privation affected the reaction of all subjects negatively, a result that cannot be explained by the criticism just cited.

The second variable relevant to this issue asked subjects: "How did you feel about what you said on tape? To what extent did it represent your views?" If, in accordance with the alternate explanation, dissonance already had been dissipated among subjects receiving low pay prior to making the discrepant statements, then the latter should have assessed the statements as compatible with their views, and the opposite should have occurred among those receiving high pay. However, once again the data do not conform to such a pattern but instead yield an interaction between privation and anality, significant beyond the .01 level (see Table 6), thus paralleling results for the variable testing boredom of tasks.

It should be freely acknowledged that results might have been more clear-cut had the reward been more definitely linked to the discrepant statements, but it would seem that the dissonant situation[11] embraces the entire complex of tasks *with* statements, not statements alone, and that the data support the hypotheses suggested in this paper.

Perhaps an even more serious deviation from the original model involves the fact that the subjects made discrepant statements on a tape recorder to be played back to the incoming subject instead of attempting to convince the subject by direct confrontation. This poses the issue of whether or not dissonance was aroused. Unfortunately, no objective methods have been stipulated for the detection of dissonance or its measurement. It seems clear that the dissonance manipulation in the current experiment was more attenuated than in the original. However, the postexperimental questionnaire revealed that in general subjects found the discrepant statements dishonest, and experienced free choice in committing themselves to making these statements—the major general specifications for dissonance arousal (Brehm & Cohen, 1962). Moreover, all groups found the tasks at least mildly boring yet stated on tape the experiment was "exciting." This fits the Festinger (1957) definition of dissonance, which states, "Two elements are in a dissonant relation if . . . the obverse of one element would follow from another [p. 13]." It also fits Festinger's criterion of nonconformity with group standards, since the subjects deceived a peer, thus violating a basic ethic of our culture.

A shortcoming of both experiments is that the tasks were not as boring (particularly to the high anals) as might have been desirable for arousal of maximal dissonance effects. During the pilot study, students referred to these tasks as "Chinese torture," and made comments such as, "Will this never end?" The fact that so many subjects nevertheless rated these tasks close to the neutral point may be related to Orne's (1962) finding that subjects tend to insert meaning into the most purposeless experiments, even in the face of colossal efforts on the part of the investigators to convince them otherwise.

Still another shortcoming of both experiments is that the dissonance manipulation contains consonant as well as dissonant elements. When offered a choice about making the discrepant statements, the subject is faced

[11] Data from the postexperimental questionnaire support the notion that attitudes toward the tasks persisted into the postexperimental period until the subjects were dehoaxed.

with a dilemma. Should he square with the "incoming" subject or please the experimenter? This was an area pinpointed earlier as exemplifying the variable of ambiguity. The conflict remains unverbalized, and whichever way the subject's decision falls, it can be justified.

Were the high-anal, highly deprived subjects so swamped with disappointment that there was no room for cognitive dissonance? We saw, earlier, that disappointment in relinquishing coupons was significantly more intense among the high anals than among the low. However, from the viewpoint of dissonance theory, a negative answer must be given to the question because experimentation has shown (Brehm & Cohen, 1959) that chance deprivation should enhance the magnitude of dissonance. Thus, the privation manipulation should have favored, not hindered, Festinger's theory, and efforts to reduce discomfort should have occasioned positive, not negative changes in attitude. Results for the high-anal group negate this expectation.

Did the high-anal subjects select alternate modes of dissonance reduction, such as derogation of the experiment, volunteering to get class credit, participating for the sake of science, under- or overevaluating the experiment? No significant differences prevailed for these questions between experimental groups or across personality types. There remains the possibility that low anals under high privation discharged dissonance through a positive shift in attitude, and high anals, by a negative shift. That is, one could postulate a boomerang effect. However, such an explanation places the theorist in the untenable position of predicting two opposing reactions without specifying the antecedent conditions responsible for such divergencies.

The a priori predictions in this paper, based on individual differences, seem to offer a more compelling account of the data. Described as stubborn, the high anals under high privation took a stand squarely in opposition to that sponsored by the experimenter (outer reference point); described as orderly in thought, they tended to meet ambiguity with clarity by appraising the discrepant statements as more dishonest than other groups; described as parsimonious, they responded with intense negative reactions to high privation and with relative positive reactions to low privation.

So much for reactions of the high-anal groups. What implications can be adduced from reactions of the low-anal groups? The trend of their responses fell in the direction predicted by Festinger and Carlsmith (1959), but was not significant.

A likely Freudian explanation is that, for the low anals, a reaction formation from the anal stage occurred, manifesting itself in a lack of interest in money. Hence, those low in anality would have negative reactions in high-reward conditions, positive reactions in low-reward conditions. While this interpretation conforms to the data and might be operative in the present experiment, it remains on unsure footing because it cannot account for dissonance results in experiments where subjects are not preselected.

A key concept from the Rosenberg (1965) study that might find application here is the subject's tendency to restructure the experiment in accordance with his subjective understanding. Perhaps such restructuring is most likely to occur where the value of the experiment is unclear. Thus, in the current study, one subject, before being dehoaxed, exclaimed,

> My parents won't believe it when I go home and tell them I spent the afternoon at the University playing with blocks. . . . Perhaps I'm on *Candid Camera!*

This comment brings to the fore the subject's search for a hidden purpose in the experiment (Orne, 1962). If the subject is geared toward finding meaning where none is obvious, then the less the overt justification for incongruent behavior, the more the subject is impelled to find *other justifications* that render his conduct reasonable. Thus, where attempts at "subjective justification" remain uncontrolled in low-justification conditions, they may induce a wide range of phenomena, such as rationalization, denial, repression, distortion, projection, or sublimation (high personal commitment). Such hidden phenomena may lead to unexpected results, as in the Brehm, Back, and Bogdonoff (1964) hunger experiments, or may explain a wide variety of paradoxical effects.

The point is that the subject may be responding to variables not made explicit within the context of the experiment, such as the hidden variables in the present study, and not made explicit by dissonance theory. In fact his behavior may be better elucidated by recourse to psychoanalytic theory since the subject seems to resort to denial, repression, reaction formation, sublimation, and similar psychological maneuvers. Certainly, where personality issues are at stake, different character types may respond differentially to the same or different variables.

Certainly, also, the less clear the demand "cue," as in the meaningless, boring tasks, and the more at odds with social norms, the more the subject will restructure the experiment subjectively; the less well defined the experimenter-subject relationship, the more the subject is likely to fall back on his usual repertoire of responding to adults in a position of authority.

This leads to a different, more extended hypothesis than that presented by Rosenberg. For, in the ambiguous circumstances just described, the best predictor of the subject's behavior derives from the assumption that he will transfer to the experimenter the role relationships expected of him by the significant persons in his early life. Thus, residues of prior anticipations and attitudes may spill over as transference effects to current situations involving perceived authority. Where there is mild stress, usually also there is mild regression (Freud, 1920). Where there is mild regression, greater dependency emerges, and a stronger pull is exerted for the arousal of transference effects. Forced-compliance experiments fall into this paradigm, in view of the difficult, or, in the present case, counternorm quality of the request made by the experimenter, designed to induce the subject to make misleading statements that will deceive a peer. In such situations, transference effects should readily surface.

If this be the case, then a transference hypothesis would be more parsimonious than the dissonance-theory hypothesis since it could explain the disparate responses of all four experimental groups. Clearly, the evidence in this paper regarding the high-anal character rests on a transference model, since the genetic theory of the anal character develops from interpersonal interactions between the child and the mothering one at a very early stage in life and presupposes the transference of early reactions to other situations and other persons in life, thus establishing the foundation for a later stable character pattern.

As for the low anals, diary notes taken during the period of postexperimental questioning in the current study indicate that they tended to denigrate compensation for participation in the experiment, lending support either to the Rosenberg or the reaction-formation hypothesis. It is suggested that both hypotheses may stem genetically from transference effects since both presume that subjects are responding to the experimenter in accordance with their own individual patterns and interaction systems in relation to a symbolic authority figure. However, the interpersonal transference hypothesis has not been submitted per se to empirical test.

Some authors (Bergin, 1962; Davids, 1964; Festinger & Bramel, 1962; Levy, 1963) have attempted to relate dissonance theory to psychotherapy. This study suggests caution in making highly discrepant interpretations to high anals who might react to such interpretations with entrenched resistance and thus incur therapeutic failure. Levy (1963), in associating dissonance with psychotherapy, gave considered weight to such potential pitfalls.

The findings tend to qualify the universality of dissonance theory and to strengthen the notion that a plurality of simultaneously operative "silent variables" govern the behavior of subjects in forced-compliance situations. Since these variables have remained unaccounted for in preceding dissonance experiments, the question is raised as to whether results attributed to dissonance formulations in the past might not stem, at least in part, from reactions to these or other unrecognized variables.

REFERENCES

ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men.* Pittsburgh: Carnegie Press, 1951. Pp. 177–190.

BERGIN, A. E. The effect of dissonant persuasive communications upon changes in a self-referring attitude. *Journal of Personality,* 1962, **30**, 423–437.

BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *Journal of Abnormal and Social Psychology,* 1959, **58**, 383–387.

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance.* New York: Wiley, 1962.

BREHM, M. L., BACK, K. W., & BOGDONOFF, M. D. A psychological effect of cognitive dissonance under stress and deprivation. *Journal of Abnormal and Social Psychology,* 1964, **69**, 303–310.

COUCH, A., & KENISTON, K. Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology,* 1960, **60**, 151–174.

DAVIDS, A. The relation of cognitive dissonance theory to an aspect of psychotherapeutic practice. *American Psychologist,* 1964, **19**, 329–332.

DEUTSCH, M., & GERARD, H. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology,* 1955, **51**, 17–27.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L., & BRAMEL, D. The reactions of humans to cognitive dissonance. In A. J. Bachrach (Ed.), *The experimental foundations of clinical psychology.* New York: Basic Books, 1962. Pp. 254–279.

FESTINGER, L., & CARLSMITH, J. Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology,* 1959, **58**, 203–210.

FINNEY, J. C. Maternal influences on anal or compulsive character in children. *Journal of Genetic Psychology,* 1963, **103**, 351–367.

FISHER, S., & LUBIN, A. Distance as a determinant of influence in a two-person serial interaction situation. *Journal of Abnormal and Social Psychology,* 1958, **56**, 230–238.

FREUD, S. *Introductory lectures to psychoanalysis.* (Delivered 1915–1916) New York: Boni & Liveright, 1920.

FREUD, S. Three contributions to the theory of sex. (Orig. publ. 1905) (Ed. & trans. by A. A. Brill) In, *The basic writings of Sigmund Freud.* New York: Modern Library, 1938. Pp. 553–639.

FREUD, S. Character and anal erotism. (Orig. publ 1908) In, *Collected papers.* Vol. 2. New York: Basic Books, 1959. Pp. 45–50. (a)

FREUD, S. On the transformation of instincts with special reference to anal erotism. (Orig. publ. 1916) In, *Collected papers.* Vol. 2. New York: Basic Books, 1959. Pp. 164–171. (b)

GRYGIER, T. G. *Dynamic Personality Inventory.* London: National Foundation for Educational Research in England and Wales, 1956.

JANIS, I. L., & GILMORE, J. B. The influence of incentive conditions on the success of role-playing in modifying attitudes. *Journal of Personality and Social Psychology,* 1965, **1**, 17–28.

JONES, E. *Papers on psycho-analysis.* (Orig. publ. 1912) Boston: Beacon Press, 1961.

KRAMER, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics,* 1956, **12**, 307–310.

LEVY, L. *Psychological interpretation.* New York: Holt, Rinehart & Winston, 1963.

ORNE, M. T. On the social psychology of the psychological experiment. *American Psychologist,* 1962, **17**, 776–783.

RICKERS-OVSIANKINA, M. A. *Rorschach psychology.* New York: Wiley, 1960.

RORSCHACH, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.

ROSENBERG, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology,* 1965, **1**, 28–43.

SNEDECOR, G. W. *Statistical methods applied to experiments in agriculture and biology* (5th ed.) Ames: Iowa State College Press, 1956.

ZIMBARDO, P. G. Involvement and communication discrepancy as determinants of opinion conformity. *Journal of Abnormal and Social Psychology,* 1960, **60**, 86–94.

(Received June 29, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 37–46

LEVEL OF ASPIRATION AND PERFORMANCE VARIABILITY

N. T. FEATHER

University of New England, Australia

Male and female Ss were tested individually in a traditional level of aspiration situation using a task that allowed E complete control over S's performance. There were 3 different conditions of performance variability: high, medium, and low. Measures of n Achievement, debilitating anxiety, facilitating anxiety, intolerance of ambiguity, and category width were available. A number of aspiration measures were employed. Results suggest that under the condition of high performance variability, Ss in whom by assumption the motive to achieve success is stronger than the motive to avoid failure ($M_S > M_{AF}$) tend to set higher aspirations, to show smaller shifts in levels of aspiration following success and failure, and to make fewer atypical responses than Ss in whom by assumption $M_{AF} > M_S$. There are no significant relationships between aspiration measures and personality variables in the low-performance-variability condition. Ss tend to be less responsive to success and failure and to make more atypical responses in the low-variability condition. It is concluded that relationships between aspiration measures and personality variables are more likely to appear under high-variability conditions. Where performance variability is low, aspiration is more in the nature of a cognitive judgment determined by performance.

An important variable in performance situations is the degree to which scores fluctuate from trial to trial. At one extreme, performance scores may show a fairly orderly change as a person works at a task. Such low variability would occur, for example, in situations that involve the learning of a skill. At the other extreme, performance scores may fluctuate in an apparently haphazard manner from trial to trial as would occur, for example, if scores were determined by a random device. One would expect differences in performance variability to influence goal-setting behavior and, indeed, there is evidence that this is so. Sutcliffe (1955b) investigated the influence of task variability on the level of aspiration in an intensive study concerned mainly with the generality of aspiration behavior across tasks. Using a set of tasks differing in performance variability, he discovered that aspirations were dependent upon performance in low-variability tasks, but became increasingly independent of performance as one moved through medium-variability to high-variability tasks. Furthermore, when the effects of performance were partialed out, aspirations showed greater generality on high- and medium-variability tasks than on low-variability tasks. Sutcliffe (1955b) interpreted these results as indicating,

> . . . that aspiration in a "clearly structured" situation amounts to a judgment or reflection of the objective situation; while aspiration in an ambiguous situation reveals the operation of personality traits [p. 72].

Generality (or similarity) in goal-setting behavior on high-variability tasks would mainly be determined by relatively stable personality differences since actual performance would not be of much help in estimating future performance. In a parallel paper, Sutcliffe (1955a) suggested that the personality trait measured by high-variability tasks might be lability of goal setting:

> Success and failure accordingly operate to give direction to the factor, producing in the first place optimism (and hence elevation of the level of aspiration) and in the second pessimism (and hence depression of the level of aspiration) [pp. 43–44].

Subsequently he showed that for tasks where performance scores were in fact determined by chance, subjects' belief in control biased their aspirations (Sutcliffe, 1956). Aspirations were higher when subjects believed that they were responsible for their performance than when their scores were apparently determined by chance.

The present investigation was designed in an attempt to replicate some of Sutcliffe's findings concerning the effects of performance variability on measures derived from the traditional level of aspiration situation. The study involved detailed analysis of a number of different aspiration measures. In addition, however, it included a number of personality variables to test the general prediction that the relationship between such variables and level of aspiration measures would be more apparent in a situation involving high performance variability than in one involving low performance variability. Choice of personality variables was guided by prior research. Some had been used in previous studies involving achievement situations (viz., n Achievement, debilitating anxiety, facilitating anxiety); others were related to the ambiguous quality of the high performance-variability task (intolerance of ambiguity) and the types of estimates subjects were required to provide (category width).

METHOD

Subjects

The subjects were 26 male and 80 female first-year undergraduates at the University of New England in 1965. The personality data were collected in group sessions; subjects were individually tested in a level of aspiration situation approximately 2 months later.

Personality Variables

The following tests were administered:

1. The projective measure of n Achievement. Subjects wrote stories to four pictures (numbers 2, 48, 1, and 7 from Atkinson's, 1958, list) under neutral conditions of administration. Protocols were scored by an experienced scorer.[1]
2. The Achievement Anxiety Test (Alpert & Haber, 1960). Scores for debilitating anxiety and facilitating anxiety were obtained from this test.
3. A test of intolerance of ambiguity (Budner, 1962), assumed to measure the degree to which a person perceives ambiguous situations as sources of threat and unambiguous situations as desirable.
4. A test of category width (Pettigrew, 1958), assumed to measure individual consistency in the range or width of the cognitive categories used by persons.

Measurement of Level of Aspiration

Task. A task was used which permitted the experimenter to report prearranged performance scores in a situation where the subject would see his performance as due to his own efforts. The task was a slightly improved version of apparatus developed by Sky (1950). It consisted of an upright stand attached to a solid base and had a vertical groove which held a cradle. The subject could raise this cradle by means of a cord which was attached to the cradle and which passed over a pulley at the top of the stand. A cord attached to the base of the cradle passed to a counterweight to give stability.

The top surface of the cradle was conical in shape and truncated. A ⅜-inch steel ball bearing was balanced upon the frustrum of the truncated cone. A vertical scale marked off in ¼ inches to give a centigrade scale (from 0 to 100) was attached to the upright stand and faced the subject. A similar scale on the back of the stand faced the experimenter. A pointer fixed to the cradle gave readings to both the subject and the experimenter.

The ball resting upon the cradle was held in position using a concealed electromagnet. It could be released by cutting off the current by means of a silently operated foot switch concealed from the subject. The ball would then roll off the cradle into a small, sloping trough attached to the cradle. The experimenter therefore had complete control over the ball. He could release it from the cradle at whatever score he wished the subject to obtain and without the subject's knowledge.

When the apparatus was not in use the cradle rested so that the pointer was level with the zero mark on each scale. The apparatus was set up on a desk and wires to batteries, etc., were concealed from the subject's view.

Procedure. Instructions were as follows:

> This is a test of manual dexterity so try to do your very best. Your task is to pull this string very gently and see how far you can raise this cradle without upsetting the steel ball which will be placed on the cradle. You will be given a number of trials. On each trial the score that you obtain depends upon how high you can raise the cradle without upsetting the ball. Your score is read from the scale. Now just try raising the cradle so that you get some idea of the tension which must be applied to raise the cradle gently. Be sure to raise the cradle gently. [The subject was given one practice trial without the steel ball in place.]
>
> Now before you begin the test proper I want you to tell me first of all what is the highest score you think you might possibly get on the first trial. Now what is the lowest score you think you might possibly get on the first trial? Now what is the score that you are actually going to try to get on the first trial?

After each trial the experimenter reported to the subject the score obtained. The experimenter could read this score from the scale facing him. The subject could also see that this was his score by examining the corresponding scale in front of him. In fact, however, the score obtained on each trial was completely controlled by the experimenter as explained above. The subject was then asked to state his highest score expected, lowest score expected, and aspiration for the next trial. These three questions were not asked at the end of the eleventh trial which was the last trial.

---

[1] I am indebted to Margaret R. Saville for scoring this test.

Performance sequences. The following performance sequences were approximated in each condition: *low performance variability,* 30, 31, 34, 38, 43, 51, 58, 63, 65, 67, 70; *medium performance variability,* 30, 34, 31, 38, 43, 58, 51, 63, 67, 65, 70; *high performance variability,* 30, 65, 58, 43, 67, 31, 51, 63, 38, 34, 70. It can be seen that each sequence involved the same set of numbers but in different order, with a mean of 50 and a range from 30 to 70. The experimenter tried to approximate the sequence in any given condition as closely as possible, making allowance for subjects who pulled the string quickly, etc. The reported performance sequences departed only minimally from those listed above. Twenty-six female subjects were tested in the low-variability condition, 24 in the medium-variability condition, and 30 in the high-variability condition. Due to the small number of male subjects available, the medium-variability condition was excluded for these subjects. Twelve male subjects were tested in the low-variability condition, and 14 in the high-variability condition.

Level of aspiration scores. The following scores were obtained for each subject: (*a*) mean highest score expected, (*b*) mean lowest score expected, (*c*) mean goal discrepancy, (*d*) mean attainment discrepancy, (*e*) difference between levels of aspiration at Trial 11 and Trial 1, (*f*) mean responsiveness after success, (g) mean responsiveness after failure, (*h*) number of atypical responses.

The discrepancy scores were defined in the usual manner (Lewin, Dembo, Festinger, & Sears, 1944; Sutcliffe, 1955b); that is, a goal-discrepancy score was the difference between the level of aspiration at Trial $(n+1)$ and performance at Trial n and was taken as positive if aspiration exceeded performance; an attainment-discrepancy score was the difference between the level of aspiration at Trial n and performance at Trial n and was taken as positive if performance exceeded aspiration. Means of the 10 goal-discrepancy scores and 11 attainment-discrepancy scores were computed for each subject.

The responsiveness scores were the same as those used by Sutcliffe (1955a). Mean responsiveness after success indicated the mean change in level of aspiration following success taking account of both sign and size of change. Increases in level of aspiration were taken as positive, decreases as negative. Mean responsiveness after failure similarly indicated the mean change in level of aspiration following failure. In the usual manner a success was defined as occurring when a subject attained or exceeded his level of aspiration on a given trial, a failure when he did not attain his level of aspiration. Number of atypical responses was defined as the number of times a subject raised his level of aspiration after failure and lowered his level of aspiration after success.

RESULTS AND DISCUSSION

Effects of Performance Variability

Table 1 presents means and standard deviations of the various level of aspiration measures listed above, for male and female subjects in relation to the variability conditions. Figures 1, 2, and 3 present mean scores over trials for female subjects in relation to the variability conditions. The mean scores over trials for male subjects are very similar to those in Figures 1 and 3 and are therefore not presented. The means for each of the aspiration measures discussed below do not differ significantly between males and females. Since the pattern of results is very similar across sex, results for males and females are discussed together and differences between males and females noted where they occur.

Mean highest expected score. Analyses of variance of subjects' highest expected scores in relation to trials and variability conditions indicate that the means differ significantly over conditions (males, $F = 5.99$, $df = 1/24$, $p < .001$; females, $F = 4.02$, $df = 2/77$, $p < .05$) and over trials (males, $F = 23.54$, $df = 10/240$, $p < .001$; females, $F = 81.15$, $df = 10/770$, $p < .001$). The mean highest expected score decreases as one moves from the high-variability condition, through the medium-variability condition, to the low-variability condition (see Table 1). This could be due to the particular performance sequence used in the high-variability condition where a high performance score (65) was reported to subjects on Trial 2, thereby indicating at an early stage the possibility of obtaining high scores on the task. The interactive effect of trials and variability condition is also highly significant (males, $F = 9.04$, $df = 10/240$, $p < .001$; females, $F = 18.95$, $df = 20/770$, $p < .001$). It is obvious from Figures 1, 2, and 3 that the means show a fairly regular increase over trials in the low-variability condition and medium-variability condition but not in the high-variability condition where changes in the means over trials are quite irregular. Means in all conditions, however, appear to follow the performance sequence. An increase/decrease in mean performance from Trial n to Trial $(n+1)$ is followed by an increase/decrease in mean

TABLE 1

MEANS AND STANDARD DEVIATIONS OF LEVEL OF ASPIRATION MEASURES FOR DIFFERENT CONDITIONS OF PERFORMANCE VARIABILITY

Condition	Males		Females		
	Performance variability		Performance variability		
	High ($N = 14$)	Low ($N = 12$)	High ($N = 30$)	Medium ($N = 24$)	Low ($N = 26$)
Mean highest score expected					
M	64.55	56.73	60.74	56.51	54.69
SD	8.89	6.30	8.52	8.68	6.88
Mean lowest score expected					
M	34.01	31.86	29.30	30.60	33.11
SD	8.62	11.13	12.21	10.76	10.95
Mean aspiration level					
M	73.58	63.89	66.66	60.14	65.34
SD	20.66	15.27	15.69	16.27	18.12
Mean goal-discrepancy score					
M	26.04	16.38	20.03	12.44	17.95
SD	20.89	14.71	15.45	15.88	17.76
Mean attainment-discrepancy score					
M	−23.39	−13.75	−16.64	−10.08	−15.16
SD	20.59	15.23	15.64	16.25	18.26
Aspiration difference[a]					
M	0.36	21.08	14.77	16.83	22.65
SD	23.85	26.21	26.83	27.48	27.15
Mean responsiveness to success					
M	8.36	6.45	14.77	7.08	5.34
SD	7.67	5.45	13.34	5.25	4.21
Mean responsiveness to failure					
M	− 3.30	− 1.67	− 3.25	− 2.44	− 1.27
SD	6.44	5.29	4.09	6.36	6.95
No. atypical responses					
M	1.00	2.75	.93	1.50	1.65
SD	.93	2.17	1.00	1.44	1.77

[a] Trial 11–Trial 1.

highest expected score from Trial $(n + 1)$ to Trial $(n + 2)$. Trial-by-trial changes in performance therefore influence trial-by-trial changes in highest expected scores in all conditions.

Mean lowest expected score. Means do not differ significantly over conditions. This could be due to the fact that all subjects had the same low performance score (30) reported to them at the outset, and subsequent performance did not fall below this score. Means differ significantly over trials (males, $F = 29.60$, $df = 10/240$, $p < .001$; females, $F = 83.88$, $df = 10/770$; $p < .001$), and the in-

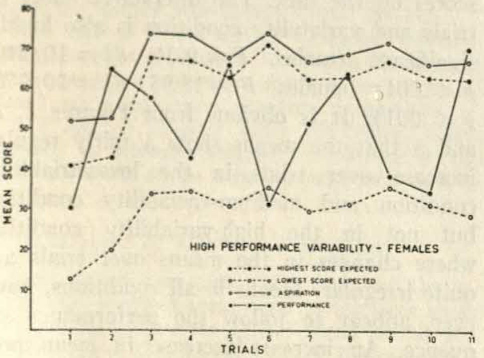


FIG. 1. Mean scores in high-performance-variability condition for female subjects.

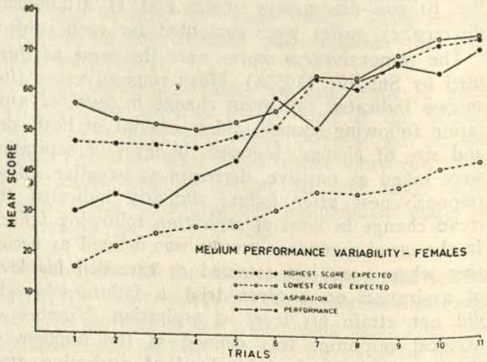


FIG. 2. Mean scores in medium-performance-variablity condition for female subjects.

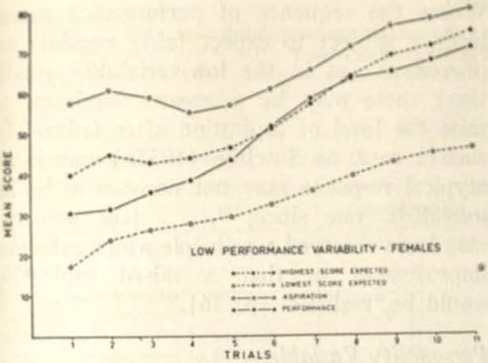


Fig. 3. Mean scores in low-performance-variability condition for female subjects.

teraction of trials and conditions is significant (males, $F = 13.38$, $df = 10/240$; $p < .001$; females, $F = 11.09$, $df = 20/770$, $p < .001$). The basis for this interaction is again seen in the less regular trial-by-trial changes in the means in the high-variability condition (see Figure 1). It is also apparent that changes in the means follow preceding changes in performance in all conditions.

Mean aspiration. Means do not differ significantly over conditions but they do differ significantly over trials (males, $F = 4.31$, $df = 10/240$, $p < .001$; females, $F = 31.69$, $df = 10/770$, $p < .001$). As in the preceding analyses, the interaction of trials and conditions is significant (males, $F = 4.71$, $df = 10/240$, $p < .001$; females, $F = 12.00$, $df = 20/770$, $p < .001$). Trial-by-trial changes in means are more irregular in the high-variability condition than in the other two conditions. But again means tend to follow preceding performance changes. It is interesting to note that mean aspirations exceed mean highest expected scores over trials (see Figures 1, 2, and 3). Presumably this difference reflects the difference in question asked. The question eliciting aspirations emphasized "trying" and that eliciting highest expected scores emphasized "expectations." Previous studies have found similar differences in scores depending upon the question asked (Lewin et al., 1944).

Mean discrepancy scores. Analyses of variance of subjects' mean goal-discrepancy scores and mean attainment-discrepancy scores reveal no significant differences in the means of these scores across variability conditions for either male or female subjects. As in Sutcliffe's (1955b) investigation there is a tendency for aspirations to be higher relative to performance in the high-variability condition (see Table 1), but the F ratios are not significant.

Aspiration difference (Trial 11–Trial 1). Analyses of variance of subjects' difference scores indicate that the means do not differ significantly across variability conditions for either male or female subjects. This rather crude difference measure has sometimes been employed as a general measure of responsiveness (Sutcliffe, 1955b). However, more information is given by separate measures of responsiveness that take account of both degree and direction of changes in levels of aspiration following success and failure.

Responsiveness to success and failure. Table 1 indicates that the means of subjects' mean responsiveness to success scores are in all cases ***positive*** and that the corresponding means of subjects' mean responsiveness to failure scores are in all cases ***negative***. Analyses of variance show highly significant differences between the overall mean of the success-responsiveness scores and the overall mean of the failure-responsiveness scores (males, $F = 22.75$, $df = 1/24$; $p < .001$; females, $F = 89.94$, $df = 1/77$, $p < .001$). These results replicate the frequently observed tendency (Lewin et al., 1944) for subjects to raise aspirations after success (positive shift) and to lower them after failure (negative shift). For female subjects the overall mean amount of change following success or failure decreases as one moves from the high-variability condition through the medium-variability condition, to the low-variability condition ($F = 4.20$, $df = 2/77$, $p < .05$), but these means do not differ significantly for male subjects. For female subjects also, the interaction of success versus failure and variability conditions is significant ($F = 8.07$, $df = 2/77$, $p < .01$). Table 1 indicates that for female subjects the mean of the responsiveness to success scores shows increasing divergence from the mean of the responsiveness to failure scores as one moves through the variability conditions from low, to medium, to high. This interactive effect does not occur for male subjects. This trend towards divergence of the responsiveness scores also occurred in Sutcliffe's (1955a) investigation. He accounted for it in terms of the difference in variability of the performance sequences.

When we examine each performance sequence separately and partition positive changes in performance from trial to trial (i.e., performance increases) from negative changes in performance from trial to trial (i.e., performance decreases), the mean *positive* change in performance for female subjects is 25.17 for the high-variability condition, 7.49 for the medium-variability condition, and 4.00 for the low-variability condition. For male subjects these means are 25.59 (high-variability condition) and 4.04 (low-variability condition). The mean *negative* change in performance from trial to trial for female subjects is −17.35 for the high-variability condition, −4.04 for the medium-variability condition, and 0 for the low-variability condition. For male subjects these means are −17.74 (high-variability condition) and 0 (low-variability condition).[2] The increasing divergence noted above in the means of success-responsiveness and failure-responsiveness scores with increase in performance variability is consistent with increasing divergence in the average positive and average negative change in performance as performance variability increases.

Number of atypical responses. Mean number of atypical responses increases for both male and female subjects as variability of performance decreases (see Table 1). The difference in means is significant for males ($F = 6.96$, $df = 1/24$, $p < .05$) but not for females. Almost all of the atypical responses recorded were of the same type, namely, raising the level of aspiration after failure. The other type of atypical response—lowering the level of aspiration after success—was extremely rare. These results suggest that prevalence of atypical responses may be a function of the type of performance sequence. Where the sequence of performance scores leads a subject to expect fairly regular improvement (as in the low-variability condition) there may be a greater tendency to raise the level of aspiration after failure. In such a case, as Sutcliffe (1955a) noted, an atypical response may not necessarily be an *unrealistic* one since, ". . . [the subject] may have observed a principle which promises improvement so that a raised aspiration would be 'realistic' [p. 36]."

Personality Variables

With the exception of n Achievement for males ($F = 5.44$, $df = 1/24$, $p < .05$), means of the personality variables do not differ significantly across the different conditions of performance variability. Nor are there any significant differences in the means of personality variables between males and females. In particular, there is no statistically significant evidence that males score higher on category width than do females although the means are in that direction. This lack of significance is consistent with that of another study using a different group of Australian students (Feather, 1967), but does not agree with Pettigrew's (1958) finding of higher category-width scores for males, using American students. The higher n Achievement scores among males in the high-variability condition of the present study appear to be an accident of sampling.

Intercorrelation of Variables

In view of the general lack of significant differences in means between male and female subjects and in order to augment the Ns, the data for these subjects were combined and intercorrelations involving aspiration measures and personality variables were calculated separately for the high-performance-variability condition and for the low-performance-variability condition using the combined male and female groups. Table 2 presents correlations between aspiration measures and n Achievement and debilitating anxiety scores for the two performance conditions.[3]

---

[2] The slight differences in means between males and females for each performance sequence are due to the fact that the three performance sequences were approximated as explained in the Method section.

[3] Facilitating anxiety, intolerance of ambiguity, and category width are not included as variables in Table 2 since they show insignificant correlations with aspiration measures in both performance con-

TABLE 2

INTERCORRELATION OF VARIABLES FOR HIGH-PERFORMANCE-VARIABILITY AND LOW-PERFORMANCE-VARIABILITY CONDITIONS[a]

Aspiration measures	Performance condition			
	High variability		Low variability	
	n Achievement	Debilitating anxiety	n Achievement	Debilitating anxiety
Mean highest score expected	−.05	−.06	.13	.10
Mean lowest score expected	−.07	−.11	.20	−.17
Mean goal-discrepancy score	−.02	−.35*	−.03	−.19
Aspiration difference[b]	−.34*	−.10	−.02	.11
Mean responsiveness to success	−.37*	.39**	−.01	.02
Mean responsiveness to failure	−.18	−.35*	.16	−.01
Number atypical responses	−.25	.37*	−.05	−.02

[a] For high performance variability, $N = 44$. For low performance variability, $N = 38$.
[b] Trial 11–Trial 1.
* $p < .05$ (2-tailed test).
** $p < .01$ (2-tailed test).

Differences across performance conditions. In Table 2 both n Achievement and debilitating anxiety enter into significant correlations with aspiration measures in the high-variability condition. But there are no significant relationships between these personality variables and aspiration measures in the low-variability condition. This confirms the general prediction of this study that such relationships would be more evident in situations where performance shows marked variation from trial to trial.

Turning to the specific results in Table 2, it can be seen that as n Achievement increases subjects tend to show smaller increases in levels of aspiration from Trial 1 to Trial 11 ($r = -.34$, $p < .05$) and smaller increases in levels of aspiration after success ($r = -.37$, $p < .05$). As debilitating anxiety increases, subjects tend to obtain lower mean goal-discrepancy scores ($r = -.35$, $p < .05$), to show both larger *increases* in levels of aspiration after success ($r = .39$, $p < .01$) and larger *decreases* after failure ($r = -.35$, $p < .05$), and to make more atypical responses ($r = .37$, $p < .05$). Apparently, therefore, subjects high in debilitating anxiety tend to set lower levels of aspiration relative to immediately preceding performance, to show larger shifts in aspiration, and to make more atypical shifts than do subjects low in debilitating anxiety. When these atypical shifts occur in the present study, they mainly involve an increase in the level of aspiration following failure. None of these tendencies is apparent among subjects in the low-variability condition (see Table 2).

The above differences are also evident in Table 3 where means are presented for subjects simultaneously classified with respect to n Achievement and debilitating anxiety. Cutting points close to the median were used for each distribution. In the high-variability condition subjects were classified as high in n Achievement if they had scores above 3 ($N = 21$) and high in debilitating anxiety if they had scores above 26 ($N = 21$). Otherwise they were classified as low in the respective variable. In the low-variability condition subjects were classified as high in n Achievement if they had scores above 2 ($N = 18$) and high in debilitating anxiety if they had scores above 26 ($N = 20$).[4] Otherwise they were classified as low. The

---

ditions. Mean attainment discrepancy is not included as a variable since it correlates −.99 with mean goal discrepancy for both performance conditions. Similarly mean aspiration level is not included as a variable since there is a perfect negative correlation between mean attainment-discrepancy scores and mean aspiration levels due to the fact that mean performance level shows negligible difference across subjects because of the controlled nature of the performance sequence.

[4] The cutting points in the n Achievement distribution vary by 1 point because a common cutting point (either 2 or 3) resulted in a fairly uneven distribution of subjects across categories. The use of slightly different cutting points overcame this problem.

proportion of females to males in each of the groups in Table 3 is roughly equivalent (2:1) and does not differ significantly across groups. Table 3 shows that in the high-variability condition low-high subjects have lower mean goal-discrepancy scores, larger mean positive-aspiration differences (Trial 11–Trial 1), larger mean positive increases in levels of aspiration following success, larger mean negative decreases in levels of aspiration following failure, and a higher mean frequency of atypical responses than do high-low subjects. These differences between high-low and low-high subjects are, in general, less apparent in the low-variability condition.

Finally, the results also indicate that subjects who show relatively large increases in levels of aspiration after success in the high-variability condition tend to make more atypical responses ($r = .33$, $p < .05$). In the low-variability condition subjects who show relatively small decreases in levels of aspiration after failure tend to make more atypical responses ($r = .33$, $p < .05$). Also in the low-variability condition subjects who show relatively small increases in levels of aspiration from Trial 1 to Trial 11 tend to obtain higher mean goal-discrepancy scores ($r = -.35$, $p < .05$).

Similarities across performance conditions. In both conditions subjects with relatively high mean goal-discrepancy scores tend to show smaller increases in levels of aspiration following success (high variability, $r = -.21$, *ns*; low variability, $r = -.65$, $p < .001$), and smaller decreases in levels of aspiration following failure (high variability, $r = .57$, $p < .001$; low variability, $r = .26$, *ns*). Also in both conditions subjects with relatively large differences in levels of aspiration between Trial 1 and Trial 11 tend to show larger increases in levels of aspiration following success (high variability, $r = .33$, $p < .05$; low variability, $r = .51$, $p < .001$) and smaller decreases in levels of aspiration following failure (high variability, $r = .61$, $p < .001$; low variability, $r = .51$, $p < .001$). These results suggest that subjects who set high levels of aspiration across trials tend to be less responsive to success and failure experiences than do subjects who show relatively large increases in levels of aspiration between Trial 1 and Trial 11. The former subjects would, of course, tend also to have higher mean goal-discrepancy scores.

Debilitating anxiety and facilitating anxiety are negatively correlated for subjects in both performance conditions (high variability, $r = -.47$, $p < .001$; low variability, $r = -.57$, $p < .001$). This negative correlation replicates previous results (e.g., Feather, 1963a). Facilitating anxiety is not, however, involved in any significant correlations with aspiration measures.

In neither performance condition is there a significant negative relationship between the mean responsiveness scores. Sutcliffe (1955a) found such a relationship in his study for a high-variability task but not for a low-variability task. He used this as the basis for postulating a general factor of "lability of goal-setting," which he assumed to operate under conditions of high performance variability, subjects showing greatest positive responsiveness to success also showing greatest negative responsiveness to failure, and subjects showing least positive responsiveness to success showing least negative responsiveness to failure.

Finally, none of the correlations involving intolerance of ambiguity and aspiration measure are significant. It was thought that some

TABLE 3

Mean Scores for Subjects Simultaneously Classified for n Achievement and Delibitating Anxiety

Condition	n Achievement	Delibitating anxiety	N	Mean goal-discrepancy score	Mean aspiration difference	Mean responsiveness to success	Mean responsiveness to failure	No. atypical responses
High performance variability	High	Low	9	29.00	11.11	4.59	−1.16	.44
	Low	High	9	17.51	23.33	21.89	−2.17	1.67
Low performance variability	High	Low	8	20.55	16.13	5.47	− .16	2.13
	Low	High	10	16.94	18.00	5.92	−1.82	2.70

significant relationship might emerge in the high-variability condition where there is ambiguity and uncertainty in the sense that it is difficult to predict future performance, but no specific predictions were made. Nor do category-width scores correlate with any of the aspiration measures. We expected a positive correlation between category-width scores and the size of the difference between the mean highest expected score and the mean lowest expected score (i.e., a range measure) across subjects. An analysis of the data shows, however, that this prediction is not supported in either of the variability conditions.[5]

CONCLUSIONS

These results are in general accord with Sutcliffe's (1955b) argument that where the level of aspiration situation is clearly structured in the sense that changes in performance from trial to trial are regular and consistent, aspiration is closely determined by performance and the situation becomes one that mainly involves cognitive judgment. Where, however, performance levels fluctuate markedly from trial to trial, the situation becomes more ambiguous, and relationships between aspiration measures and personality variables are more likely to become evident.[6] Changes in performance are important determinants of changes in aspiration in all conditions of the present study (see Figures 1, 2, and 3), even in the high-variability condition where the effect of personality variables is more evident. There is no evidence in the present study, however, for a factor involving lability of goal setting (Sutcliffe, 1955a).

How are these results to be interpreted theoretically? It should be apparent that we have made no attempt to derive them from the theory of achievement motivation (Atkinson & Feather, 1966). The traditional level of aspiration situation, as investigated in the present study, is highly complex, involving successive decisions between levels of possible future performance. These decisions are determined to a large extent by the payoffs and costs that would follow success and failure at each of these levels, and by the subjective probabilities of success and failure at these levels. The present task situation corresponds to what we have called a high-payoff–low-cost condition (Feather, 1964) since the subject is credited with his performance score irrespective of whether or not he attains or fails to attain his stated level of aspiration. The overall height of the aspiration curve in relation to the performance curve can be changed quite easily by manipulating the payoffs and costs. For example, the height of the aspiration curve tends to fall as the cost of failure is increased (Feather, 1964). It is likely that rather complex decision functions will be necessary if level of aspiration behavior is to be understood for all possible conditions of payoff and cost. The theory of achievement motivation (Atkinson & Feather, 1966) as presently formulated applies to choice situations involving different levels of difficulty (e.g., the ring-toss game in the study by Atkinson & Litwin, 1960) where success or failure is not associated with any gain or loss of points, as it is in the traditional level of aspiration situation.

Despite the complexity of the traditional level of aspiration situation, however, it is interesting to note that those personality variables that do enter into significant relationships with aspiration measures are the very ones that have been taken as measures of the motive to achieve success (M_S) and the motive to avoid failure (M_{AF}), that is, n Achievement and debilitating anxiety, respectively. The results in Table 3 suggest that when performance variability is high subjects in whom by assumption $M_S > M_{AF}$ (the high-low subjects) tend to set higher levels of aspiration, to show smaller shifts in aspiration following success and failure, and to make fewer atypical responses than do subjects in whom by assumption $M_{AF} > M_S$ (the low-high subjects). These results are consistent with previous findings since it is a well-documented fact that subjects high in anxiety tend to be more variable in their goal-setting behavior and to be more prone to make atypical responses (Atkinson, 1957; Moulton, 1965). The results also indicate, however, that degree of responsiveness and the tendency to make atypical responses are not unrelated to performance variability. Responsiveness tends to be greater under high-variability conditions, and the tendency to make atypical responses (especially raising aspirations after failure) is more likely to occur in low-variability conditions where performance improves from trial to trial. It is likely that the other type of atypical response (lowering aspirations after success) would be most evident under conditions where performance shows a regular decrease from trial to trial. In the traditional level of aspiration situation, therefore, neither the degree of responsiveness nor the tendency to make atypical responses can be divorced from the nature of the performance sequence, but these measures are related to personality variables when task performance is highly variable and difficult to predict.

---

[5] Category-width scores and facilitating anxiety scores are negatively correlated in the high-variability condition ($r = -.39$, $p < .01$), but not in the low-variability condition ($r = -.05$, *ns*). This result needs to be replicated. It was not predicted.

[6] We have previously found that judgments of *expectations* of success are more likely to be related to measures of basic motives when the situation is relatively unstructured (Feather, 1963b, 1965), but that once a person becomes involved in a task his performance assumes the dominant role in influencing his expectations.

REFERENCES

ALPERT, R., & HABER, R. N. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology,* 1960, **61**, 207–215.

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychological Review,* 1957, **64**, 359–372.

ATKINSON, J. W. *Motives in fantasy, action, and society.* Princeton, N. J.: Van Nostrand, 1958.

ATKINSON, J. W., & FEATHER, N. T. (Eds.) *A theory of achievement motivation.* New York: Wiley, 1966.

ATKINSON, J. W., & LITWIN, G. H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *Journal of Abnormal and Social Psychology,* 1960, **60**, 52–63.

BUDNER, S. Intolerance of ambiguity as a personality variable. *Journal of Personality,* 1962, **30**, 29–50.

FEATHER, N. T. The effect of differential failure on expectation of success, reported anxiety, and response uncertainty. *Journal of Personality,* 1963, **31**, 289–312. (a)

FEATHER, N. T. The relationship of expectation of success to reported probability, task structure, and achievement-related motivation. *Journal of Abnormal and Social Psychology,* 1963, **66**, 231–238. (b)

FEATHER, N. T. Level of aspiration behaviour in relation to payoffs and costs following success and failure. *Australian Journal of Psychology,* 1964, **16**, 175–184.

FEATHER, N. T. The relationship of expectation of success to need achievement and test anxiety. *Journal of Personality and Social Psychology,* 1965, **1**, 118–126.

FEATHER, N. T. Evaluation of religious and neutral arguments in religious and atheist student groups. *Australian Journal of Psychology,* 1967, in press.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Pp. 338–378.

MOULTON, R. W. Effects of success and failure on level of aspiration as related to achievement motives. *Journal of Personality and Social Psychology,* 1965, **1**, 399–406.

PETTIGREW, T. F. The measurement and correlates of category width as a cognitive variable. *Journal of Personality,* 1958, **26**, 532–544.

SKY, A. W. An apparatus for a frustration task. *Australian Journal of Psychology,* 1950, **2**, 116–120.

SUTCLIFFE, J. P. Responsiveness of the level of aspiration to success and failure as a function of task variability. *Australian Journal of Psychology,* 1955, **7**, 34–44. (a)

SUTCLIFFE, J. P. Task variability and the level of aspiration. *Australian Journal of Psychology,* Monogr. Suppl., 1955, No. 2. (b)

SUTCLIFFE, J. P. Judgment of random events as a function of belief in control. *Australian Journal of Psychology,* 1956, **8**, 128–139.

(Received July 6, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 47–54

LATITUDE OF ACCEPTANCE AND ATTITUDE CHANGE:

EMPIRICAL EVIDENCE FOR A REFORMULATION[1]

ALVIN L. ATKINS
Yeshiva University
KAY K. DEAUX AND JAMES BIERI
University of Texas

The latitude of acceptance concept was defined in terms of a range criterion consisting of (a) statements indicated as acceptable by each S, and (b) the scalar distance covered by these statements. 54 Ss, reflecting pro, neutral, and anti own-attitude positions toward fraternities, judged the attitude expressed toward fraternities in a series of statements and indicated which statements they found acceptable or objectionable. Each S also judged the scale position represented in 1 of 3 possible persuasive communications. Results indicated that Ss who perceived a persuasive communication as falling within their latitude of acceptance showed significantly greater change of own position in the direction of the communication than Ss for whom the communication fell outside their latitude of acceptance.

With the study of attitude-change processes becoming a major concern in social psychological research, there has been growing recognition of the importance of developing improved techniques for the assessment of attitudes. The concept of latitude of acceptance and latitude of rejection, a device for establishing an individual's stand toward a social issue, was conceived as a technique which permitted dimensional analysis of responses to various positions toward a social object (Sherif & Hovland, 1961; Sherif, Sherif, & Nebergall, 1965). The need for such an approach stemmed from the realization that to consider a single scalar position as indicating the attitude of an individual does not adequately reflect the multifaceted nature of his reactions to the social issue in question.

Analysis of the individual's latitude of acceptance provides an opportunity to examine both the range of statements which comprise his attitude as well as the susceptibility to attitude change implied by such a range of acceptable statements. Latitude of acceptance has value, then, not only as a technique for assessment of own attitude, but also as a predictive tool in attitude change. Although not attempting a direct empirical test of degree of attitude change in relation to size of latitude of acceptance and rejection, Hovland, Harvey, and Sherif (1957) proposed that attitude change will occur only under the condition in which the communication advocating change falls *within* one's latitude of acceptance. In operationalizing the concepts of latitude of acceptance and rejection, they asked their subjects to indicate the positions on an issue that were *most* acceptable and *most* objectionable as well as those other positions that were also acceptable or objectionable. As a result, they were able to investigate item placement in relation to one's own position. It was found that, in comparison to moderate subjects, extreme subjects accepted fewer statements and rejected more statements. Hovland et al. observed that subjects with extreme positions had raised thresholds of acceptance and lowered thresholds of rejection in placing items, which resulted in their having smaller latitudes of acceptance and greater latitudes of rejection than subjects with moderate positions. Since more subjects with moderate positions changed in the direction of the position advocated in the communication, it was inferred that wide latitudes of acceptance facilitate attitude change.

An alternative approach to the prediction of attitude change derives from Festinger's (1957) dissonance theory. Using this theoretical framework, Zimbardo (1960) reasoned that opinion conformity will be greatest when a communication is widely discrepant from one's own position, presumably creating maximum dissonance. Under these conditions, change should be greatest when a communication falls within one's latitude of rejection. In his experimental design, Zimbardo provided subjects with a friend's rating of a case study which was either slightly discrepant or widely discrepant for each subject. In the latitude of acceptance condition, the statement selected as the friend's "most acceptable" statement was one scale position removed from the subject's own most acceptable rating, while the latitude of rejection condition was three scale units removed. Consistent with his predictions, Zimbardo found that significantly more change was produced by widely discrepant communications falling within subjects' latitudes of rejection than by slightly discrepant communications falling within subjects' latitudes of acceptance.

---

[1] This research was supported under Grant GS-842 from the National Science Foundation.

Considering that the predictions of Hovland et al. and of Zimbardo derive logically from their respective theoretical assumptions, these conflicting results are difficult to reconcile. Despite the findings reported by Zimbardo, the expectation that attitude change should be greater when the communication falls within one's latitude of acceptance is a reasonable one which requires further empirical investigation. However, if the concept of latitude of acceptance and rejection is to have predictive utility in attitude change, we must specify a number of additional considerations.

1. In proposing a measure of latitude of acceptance, Hovland et al. (1957) actually used a frequency criterion. That is, the *number* of statements a subject considered acceptable defined his latitude of acceptance. Since a subject who held an extreme position tended to accept fewer statements and reject more statements than a subject with a more moderate position, Hovland et al. proceeded to formulate a theoretical model of latitude of acceptance which made the size of the latitude of acceptance a function of the scale position of the subject's stand on the issue. For example, an individual at the most extreme position was expected to accept the item adjacent to his own position, to be neutral toward the next statement, and to reject all those statements beyond that point. This procedure, based solely on scale position, gives no recognition to individual differences among those subjects falling at the same scale point. Rather, it assumes that a particular position necessitates a particular range of acceptance.

2. Assuming that two individuals do, in fact, endorse the same number of items, can we infer that these items actually represent equivalent scalar distances for these individuals? We need to specify not only the number of items endorsed, but also the position on the scale where each subject placed these items.

3. Both Hovland et al. (1957) as well as Zimbardo (1960) provided subjects with communications at predetermined distances from their own positions. Such a procedure fails to take into account differences among the subjects in the way in which this communication is evaluated and judged. If we are to make predictions based upon latitudes of acceptance, it is important that we allow each subject not only to specify his own range of acceptable statements, but also to indicate where this communication falls on this same scale. If the communication is differentially perceived, then we are justified in assuming that it will be associated with differential effects in the stability or change of attitudes.

In the present study we have reconceptualized latitude of acceptance by defining it in terms of a *range* criterion consisting of the scalar distance covered by the items indicated as acceptable by each subject. Furthermore, we have analyzed the validity of such an approach for predicting attitude change.

Two basic hypotheses were advanced:

Hypothesis 1: Regardless of the actual scalar position represented by the communication, attitude change in the direction of a persuasive communication will occur when the communication is perceived to lie *within* one's latitude of acceptance.

Hypothesis 2: Drawing upon the investigations of own position as an anchor (Bieri, Atkins, Briar, Leaman, Miller, & Tripodi, 1966; Sherif & Hovland, 1961), it was hypothesized that the evaluation of a persuasive communication will be subject to judgmental distortion. In order to facilitate the acceptance of a position at some distance from one's own, it is to be expected that the communication will be judged as "really not too far from one's own," making a shift in own position a plausible outcome. Such judgmental distortion is to be found in the phenomenon of assimilation. Thus, it was hypothesized that attitude change would be associated with tendencies toward assimilative distortion of the persuasive communication.

METHOD

Subjects

As part of a larger study, male college undergraduates were given a student activities questionnaire designed to yield information on students' attitudes toward extracurricular activities. Students rated statements on a number of topics, including the degree to which they felt that fraternities were desirable on the college campus. A 15-point graphic rating scale was used in which 1 indicated very undesirable, and 15 indicated very desirable. From the responses obtained in this questionnaire, 54 subjects, divided approximately equally among pro, neutral, and anti own-attitude positions, were selected at random to participate in this study. The mean scale positions for these three groups were 13.67 (pro), 8.18 (neutral), and 2.42 (anti).

Procedure

Latitude of acceptance. Thirty-three statements, compiled originally by Segall (1959) to cover the range from extremely profraternity items to extremely antifraternity items, comprised the item pool. Initially, each subject was asked to judge each statement on a 15-point graphic scale in terms of the degree of favorableness of the attitude represented in each statement. Each subject was permitted to use any number of scale positions he wished, with any frequency.

At a second session, in conjunction with other tasks, each subject was given the same series of statements and asked to do the following: (*a*) indicate the one statement closest to his own point of view on the topic, (*b*) indicate any other acceptable statement or statements, (*c*) indicate the one most objectionable statement, and (*d*) indicate any other objectionable statement or statements.

Persuasive Communication

Following this, a communication advocating one of three possible positions (moderately pro, moderately anti, and extremely anti) was presented to each subject. Within each group of pro and anti subjects, approximately half received a moderately pro communication, while the other half received a moderately anti communication. Among the neutral subjects, approximately half received a moderately anti communication, while the other half received an extremely anti communication. A moderately pro communication is presented below as a sample:

> A statement made recently by the Committee on Undergraduate Affairs at a Major Eastern University
>
> In evaluating the role of fraternities on a college campus, it is obvious that a number of factors must be taken into account. Certainly some students can function quite independently of organized groups and do not require the reassurance that others receive from feeling that they have been accepted by their peers. For those who have developed enough of a sense of security and an ability to make friends easily, fraternities may not add very much. However, statistics indicate that most undergraduates actually do depend upon the guidance and advice of others, both in intellectual and social matters. In this regard, fraternities can provide a cushion against the new environment, especially for boys coming to the impersonal campus of a large university. While many opportunities exist for the newcomer to adjust to a college setting, the fraternity system offers a quick means of gaining the fellowship of other undergraduates who are facing similar problems at the same crucial period in their lives.

In order to determine the approximate scale position of each of these communications, a group of 20 judges drawn from the same student population, but with neutral positions toward fraternities, served as a control group for providing mean judged scale positions. The mean values for the communications were 11.3 (moderately pro), 5.0 (moderately anti), and 1.9 (extremely anti).

After reading his communication, each subject was asked to consider the position advocated in the communication and to indicate which scale position on the 15-point rating scale he had previously used best represented the stand taken in the communication. Following this, each subject again indicated how desirable he felt it was to have fraternities on the college campus, in a manner identical to that used to assess his position initially.

RESULTS

The results will be presented under two major headings—the reformulation of the latitude of acceptance, and the application of this approach to the prediction of attitude change.

Reformulation of Latitude of Acceptance

Before considering a range criterion for the analysis of latitude of acceptance, the data were examined in terms of a frequency criterion, that is, by the number of statements accepted and rejected, as well as those presumably neutral items not checked which Sherif et al. (1965) characterize as the latitude of noncommitment. In contrast to the results reported by Hovland et al. (1957), who found that extreme subjects rejected significantly more items than did more moderate subjects, the differences found in this study are in the opposite direction, although not statistically significant. In Table 1 it can also be observed that there is no tendency for neutral subjects to accept more items than do extreme subjects, as is suggested by the Sherif-Hovland model.

In applying a range criterion, the following approach was used. If a subject accepted two statements, we considered where he placed these statements on the 15-point scale provided. If he had judged one of these statements to be represented by Point 7 and the other by Point 10, then his latitude of acceptance was the range of scalar distance between these two points, that is, 3 units. If a subject accepted three statements, one represented by Point 6, another by Point 7, and the third by Point 8, then his latitude of acceptance consisted of 2 scale units and was smaller than the former. In this manner we determined individually for each subject a latitude of acceptance based on the statements which he found acceptable and the range of the scale covered by these statements.

That it is necessary to consider the latitude of acceptance for each subject individually can be seen in Table 2, where the mean range of statements accepted for all subjects grouped according to their own-attitude positions is reported. As can be observed, these means do not differ significantly from each other. Thus, it appears that although subjects do vary considerably in the range of statements they are willing to accept, this variation reflects individual differences which are not consistently related to own-attitude positions.

TABLE 1

Mean Number of Statements Accepted, Rejected, and Neutral

S's own position	N	No. accepted[a]	No. rejected[a]	No. neutral[a]
Pro	18	7.61	9.39	16.00
Neutral	17	6.53	10.18	16.29
Anti	19	6.21	9.42	17.37

[a] None of the comparisons between own-position groups for number of statements accepted, rejected, or neutral was significant.

TABLE 2

Mean Range of Statements Accepted in Relation to Own Position

Own position	N	Mean range[a]	SD
Pro	18	5.83	2.69
Neutral	17	5.71	3.83
Anti	19	6.00	2.81

[a] Differences between groups are not significant.

Application of Latitude of Acceptance to Prediction of Attitude Change

Hypothesis 1. The initial hypothesis proposed that regardless of the actual distance between the communication and one's own position, those subjects who perceive the communication as lying within their latitude of acceptable statements will be more susceptible to its influence. By considering each subject's judgment of the scale position represented by the persuasive communication, it was possible to determine if the subject perceived this communication as lying within the range of statements which he endorsed. Table 3 presents the results obtained when the mean change in own position for the group of subjects for whom the communication fell *inside* their latitude of acceptance is compared with the group for whom the communication fell *outside* their latitude of acceptance. For those subjects who judged the persuasive communication as falling inside their latitude of acceptance, the average change was 2.74 units in the direction of the communication, while the average change was .25 unit in the opposite direction for the comparison group. This difference is significant at the .01 level.

Inspection of the separate own-attitude group scores in Table 3 reveals that the overall change scores cannot be attributed to the shifts for one group alone. Within the inside latitude of acceptance condition, for example, both the neutral and anti groups demonstrated considerable positive change. It should also be pointed out that within these own-attitude groups some subjects showed a shift in attitude that at times exceeded the position advocated in the communication. An analysis of variance of the change scores for the own-attitude groups in relation to latitude of acceptance indicates that a significant difference between own-attitude groups was obtained ($F = 4.99$, $p < .05$), as well as a significant interaction between own attitude and latitude of acceptance ($F = 3.57$, $p < .05$).

On the other hand, we could attempt to predict attitude change on the basis of the *distance* between the subject's own attitude and the location of the persuasive communication, as determined by preestablished control-group ratings. Such an analysis derives from the assumption that communications which are at relatively short distances from the subjects' own positions would probably fall within their latitudes of acceptance and would, therefore, probably gain greater acceptance (e.g., pro subjects receiving a moderately pro communication as compared with pro subjects receiving a moderately anti communication). Such an assumption would accord with the Sherif and Hovland theoretical interpretation. However, as reported in Table 4, no evidence was found that subjects who received a communication closer to their own position showed any significantly greater susceptibility to change than did subjects at a greater distance from the communication. Furthermore, there is no evidence to suggest that greater distance between the subject's own position and the persuasive communication results in change of attitude as a mode of dissonance reduction, as argued by Zimbardo (1960). It seems that distance as established by predetermined scale location of the communication does not in and of itself represent the crucial factor for consideration in susceptibility to change. Rather, the subject's own perception and judgment of the location of this communication, which reflect his interpretation of the content of the communication, appear to constitute an important variable in the prediction of change.

TABLE 3

MEAN CHANGE OF OWN ATTITUDE IN RELATION TO JUDGMENT OF COMMUNICATION

	Pro	Neutral	Anti	Overall (pro, neutral, & anti combined)
Communication judged inside latitude of acceptance ($N = 27$)				
Mean change[a]	+ .89	+4.33	+3.33	+2.74
SD	2.57	2.16	2.96	2.93
	(N =9)	(N =6)	(N =12)	
Communication judged outside latitude of acceptance ($N = 27$)				
Mean change	.00	−1.82	+1.86	−0.25
SD	.87	1.78	1.57	2.07
	(N =9)	(N =11)	(N =7)	

Note.—Overall mean change: $t = 4.15$, $p < .01$.

[a] In all cases, + = a change in the direction of the communication; − = a change in the direction away from the communication.

TABLE 4

MEAN ATTITUDE CHANGE IN RELATION TO DISTANCE OF COMMUNICATION FROM OWN-ATTITUDE POSITION

Own position	Moderate distance[a]	Far distance
Pro	+ .78 ($N = 9$)[b]	+ .11 ($N = 9$)[c]
Neutral	− .62 ($N = 8$)[c]	− .33 ($N = 9$)[d]
Anti	+3.33 ($N = 9$)[c]	+2.90 ($N = 10$)[b]
Mean change	+1.23	+ .96

Note.—Moderate versus far distance: $t = .329$, *ns.*

[a] In all cases, + = a change in the direction of the communication; − = a change in the direction away from the communication.

[b] Moderately pro communication.

[c] Moderately anti communication.

[d] Extremely anti communication.

Hypothesis 2

The second hypothesis posits that own position will operate *differentially* as an anchor to distort judgment of the communication depending upon whether the communication is within or without the latitude of acceptance. That such distortions can operate in a systematic fashion so as to exaggerate or minimize perceived differences between one's own position and that advocated in the communication has been demonstrated (Bieri et al., 1966; Sherif & Hovland, 1961). However, it is possible to argue that a dual process of change is operating here, facilitated by the distortion of the message represented in the communication. Subjects who perceive the communication as falling inside their latitude of acceptance not only change in the direction of the communication but may, at the

TABLE 5

DISTORTION OF COMMUNICATION IN RELATION TO LATITUDE OF ACCEPTANCE

	Mean distortion[a]
Communication judged inside latitude of acceptance ($N = 27$)	+0.84
Communication judged outside latitude of acceptance ($N = 27$)	−0.56

Note.—Overall mean distortion: $t = 2.59$, $p < .05$.

[a] In this case, + = a distortion of the communication toward one's own position; − = a distortion of the communication away from one's own position.

same time, judge the communication to be closer to their own position than it actually is (assimilation). Change, then, comes about by changing own position toward the location of the communication while simultaneously changing the location of the communication in the direction of one's own position. By comparing each subject's judgment of the scale position of the communication in relation to the average judgment given to this communication by the control group, we can see the extent to which the communication is distorted in the direction of one's own position. Table 5 indicates that inside latitude of acceptance subjects did show significantly greater "assimilative" distortion of the communication toward their own position than did outside latitude of acceptance subjects. Similarly, rejection of the persuasive communication by outside latitude of acceptance subjects appears to be facilitated by a "contrast" distortion.

DISCUSSION

The results of this study provide further support for the theoretical significance of latitude of acceptance and for its predictive value when assessment of the latitude of acceptance is based upon a measure reflecting individual differences in the range of items accepted. The fact that no differences were found when comparisons were made between the attitude change obtained under conditions of relatively small as opposed to relatively large discrepancies between the position of the subjects and that advocated in the communication argues even more strongly for the value of the latitude of acceptance variable.

One may speculate at this point concerning a number of factors which may possibly mediate the influence of latitude of acceptance on attitude change. For example, it is conceivable that judging the communication as lying within the range of acceptable items simply reflects an inability on the part of such subjects to discriminate among different scale positions. If so, change may be a function of poor discriminability rather than latitude of acceptance. Based on several informational measures of ability to discriminate among statements representing different attitudinal positions toward fraternities (Bieri et al., 1966), comparisons between inside latitude of acceptance subjects and outside latitude of acceptance subjects were made. None of these comparisons was significant. Similarly, comparisons on attitude change between inside and outside latitude of acceptance subjects who differed on the Pettigrew Category Width Scale failed to yield any significant results.

Perhaps subjects who change are merely demonstrating a tendency to respond in a consistently positive manner toward all statements and communications provided for judgment. In this way, acceptance of the argument presented in the communication may reflect an acquiescent personality characteristic which may underlie the latitude of acceptance variable. Again, a comparison between inside latitude of acceptance subjects and outside latitude of acceptance subjects on the basis of scores derived from an acceptance of authority scale revealed no significant differences on this dimension.

Since we cannot attribute these latitude of acceptance findings to such underlying factors of personality or cognitive functioning, it is important to explore those unique components which this variable may possess. In his paper on communication and discrepancy, Zimbardo (1960) reports that those subjects who conformed accepted a wider range of statements initially than did nonconformers; that is, they had large latitudes of acceptance. This threshold of acceptance was thought to provide a useful potential indicator of susceptibility to persuasive communications. If we compare the mean range of inside latitude of acceptance subjects (7.52 units) with that of outside latitude of acceptance subjects (4.18 units), we also find that the former group has a significantly wider latitude of acceptance ($t = 4.54$, $p < .01$). This finding may support the interpretation that a wider latitude of acceptance results in a flexibility in attitudinal position which makes possible the restructuring of one's own stand in the face of a persuasive communication. One mechanism which may facilitate such a restructuring may be the tendency to perceive the persuasive communication as somewhat closer to one's own position than is actually the case. Through a process of assimilative distortion, subjects who changed apparently were able to perceive the persuasive communication as sufficiently close enough to their own position to accept some of its content. That such assimilative distortion was not solely responsible for change, however, can be inferred from additional findings concerning distortion and amount of attitude change. Subjects who showed the greatest degree of change did not differ significantly in degree of assimilation of the persuasive communication from those who changed the least.

To gain a clearer picture of the relevance of these findings to the broader context of attitude-change research, let us consider the features of this study which are either similar or dissimilar to previous research in this field.

1. The topic under study was both salient to the subjects and related to the natural college setting in which they lived. It has been noted (Sherif et al., 1965) that attitude-change research often deals with unfamiliar topics, highly technical material (scientific, medical detail), or issues unrelated to the everyday, ongoing experiences of the subjects. In the present research the issue of fraternities was highly salient on the campus and constituted an issue which drew strong response from the student body.

Similarly, it has been proposed (Sherif & Hovland, 1961; Sherif et al., 1965) that a number of previous findings of attitude change refer to situations where ego involvement was minimal. The present research suggests that attitude change in response to a persuasive communication is not limited to noninvolving issues. While no direct measure of involvement independent of extremity was obtained, it can reasonably be argued that the issue was ego involving for the subjects. Since the question of the desirability of fraternities on a college campus was a hotly debated topic on the campus at the time, it can be assumed that for many students this topic was an emotionally charged one. Because the subjects were randomly selected from the questionnaire responses (with no refusals), it can be assumed, further, that the sentiments prevailing on the campus generally were reflected in the sample used for the study.

2. The persuasive communication was attributed to a "Committee on Undergraduate Affairs at a Major Eastern University" in an attempt to give the source of the message high credibility. Consistent with previous research (Sherif & Hovland, 1961), it was to be expected that a highly credible source was an important variable in producing effective communication results. In this respect, then, this research is similar to earlier attitude-change designs.

3. Since three different messages were used, representing moderate and more extreme discrepancies between own position and that of the communication, we have demonstrated that discrepancy cannot be considered the sole or major determinant of attitude change. Further, since change was observed under two discrepancy conditions, it is clear that if distance is to be considered an important variable at all, it must be viewed in terms of its interaction with other subject and situational variables.

4. Strikingly, subjects who are extreme in own-attitude position demonstrated change as well as did the moderate ones. Based on earlier findings (Sherif & Hovland, 1961), it is to be expected that when change occurs, it is to be found among the moderate, less-committed subjects. In this research, we have found that both extreme antifraternity subjects as well as subjects moderate in attitude toward fraternities had shifts in own position in the direction of the communication. It would seem that future attitude-change research must consider parameters other than merely extremity and involvement of own attitude in predicting shifts in position.

5. Recent investigations (Sherif et al., 1965) have suggested that the latitudes of rejection and noncommitment be studied independently of the latitude of acceptance. In concentrating the present research upon a refinement of the latitude of acceptance variable, we have demonstrated that the latitude of acceptance is predictive of own-position change. Certainly studies are needed to indicate whether the latitudes of acceptance, rejection, and noncommitment are simply complementary variables or whether they differ in their predictive utility for attitude change.

6. Finally, it is to be noted again that in contrast to previous studies in which no assessment of each judge's perception of the communication was obtained, the subjects here were allowed to specify the location of the persuasive communication according to their own judgments. In this way, individual differences in evaluation and judgment have been utilized in an attempt to make the latitude of acceptance sensitive to the variety of interpretations and perceptions to be found even within a relatively homogeneous sample.

REFERENCES

BIERI, J., ATKINS, A. L., BRIAR, S., LEAMAN, R. L., MILLER, H., & TRIPODI, T. *Clinical and social judgment*. New York: Wiley, 1966.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

HOVLAND, C., HARVEY, O. J., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *Journal of Abnormal and Social Psychology,* 1957, **55**, 244–252.

SEGALL, M. The effects of attitude and experience on judgments of controversial statements. *Journal of Abnormal and Social Psychology,* 1959, **58**, 61–68.

SHERIF, M., & HOVLAND, C. *Social judgment*. New Haven: Yale University Press, 1961.

SHERIF, C., SHERIF, M., & NEBERGALL, R. *Attitude and attitude change*. Philadelphia: Saunders, 1965.

ZIMBARDO, P. G. Involvement and communication discrepancy as determinants of opinion change. *Journal of Abnormal and Social Psychology,* 1960, **60**, 86–94.

(Received July 11, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 55-63

NONCONFORMITY TO EXPERIMENTALLY INDUCED GROUP NORMS OF HIGH-STATUS VERSUS LOW-STATUS MEMBERS[1]

NORMA D. FESHBACH

Department of Education, University of California, Los Angeles

It was hypothesized that under conditions in which conformity was nonveridical and detrimental to the interests of the group, nonconforming responses would be a positive function of status as assessed by peer rankings, of experimentally induced status, and of the anticipation that one's judgments would remain private. Subgroups of high- and low-status Ss, each representing a campus fraternity, were given a judgmental task in which a conflict was established between the properties of the stimuli being judged and a fictitious group norm. It was found that, as predicted, high-status Ss were significantly less conforming than low-status Ss. A reverse trend was obtained for experimentally induced status, while the public-private variable showed a significant interaction with the other 2 conditions.

The present study is concerned with factors facilitating nonconforming behavior when conformity to group norms is nonveridical and is detrimental to the interests of the group. Specifically examined are the influence of an individual's actual status in a group, his experimentally manipulated status, and his expectation that his judgments will be made public.

The findings of a number of recent experiments indicate that nonconformity in competent, high-status group members is more acceptable to the group than the same deviant behavior in low-status group members (Hollander, 1960; Sabeth, 1964). Early studies of conformity to group norms in the context of functioning social groups appear to convey a rather different image of the correlates of conformity. These investigations have indicated that persons of higher status have stronger conformity tendencies (Kelley & Volkart, 1952; Newcomb, 1943), while low popularity is associated with nonconformity (Festinger, Schachter, & Back, 1950; Schachter, 1951).

While Kelley and Shapiro (1954, p. 668) noted that the nonconformist "appears to have little status and prestige and to find few attractions in the group," they questioned the generality of these findings and suggested that where conformity is detrimental to the success of the group, members who perceive themselves to be highly accepted by their peers will exhibit *less* conformity than members who feel little acceptance. Similarly, Hollander's (1958) concept of "idiosyncracy credit" implies that the usual sanctions for nonconformity are less likely to be evoked for highly valued group members. Conformity, then, to group norms which are valued and shared need not have the same functional significance as conformity to norms which, as in the group-pressure studies, are not adaptive. An attempt by Kelley and Shapiro (1954) to demonstrate this relationship proved to be unsuccessful, probably because the experimental induction of status significantly affected the value of group membership, the low-status subjects having a much lower and more variable evaluation of their group. The effects of valuation of group membership could obscure the effects of the status-acceptance variable since it has been shown that members who have a high valuation of their group will be more likely to conform to the norms of that group than members who have a low evaluation (Festinger et al., 1950; Kelley & Volkart, 1952).

---

[1] The present study was based upon a doctoral dissertation submitted to the Department of Psychology at the University of Pennsylvania. The author wishes to express her appreciation and gratitude to James C. Diggory, thesis chairman, for his invaluable guidance and encouragement, and to Malcolm Preston and to the late Miles Murphy for their interest and assistance.

The use of more stable, cohesive groups, in which it would be possible to vary status without substantially modifying the attractiveness of the group, should provide an appropriate situation in which to study the relationship between status and nonconformity, where nonconformity is useful to the group. The present investigation also proposes to study the effects on conformity of experimentally induced status as compared to more enduring, "real" status variations. Some of the inconsistency in the findings of studies relating status and conformity may be partially a function of the procedures by which status was defined and of the failure to control for the impact of experimentally induced status upon real status variations.

It is expected that high-status individuals will conform less because they are less likely to be censured by the group or, in Hollander's terms, they have accrued more "idiosyncracy credits" than low-status individuals. Anticipation of censure or fear of rejection for nonconformity is influenced not only by the individual's current status within the group, but also by the factors affecting the group's knowledge of his deviant performance. One would expect that nonconformity would be greater when the subject's judgments are private and unavailable to the group.

On the basis of these considerations, three hypotheses are proposed, assuming the following conditions obtain: the use of relatively cohesive groups with high valuation of the group by its members and the utilization of group norms which, if conformed to, would result in nonveridical behavior and would be detrimental to the interests of the group.

1. High-status individuals manifest significantly greater deviation from group norms than low-status individuals, where status is determined by actual peer rankings.
2. Individuals who perceive themselves to be highly accepted by their group as determined by fictitious peer rankings manifest significantly greater deviation from group norms than individuals with perceived low-group acceptance.

(As previously indicated, an interaction between status and acceptance is anticipated. Berkowitz and Macaulay's (1961) findings that individuals whose high status is insecure are highly attracted to their group suggests that the effects of acceptance may not be the same for the two status conditions.)

3. Individuals who anticipate that their judgments will be private manifest significantly greater deviation from group norms than individuals who anticipate public knowledge of their judgments.

METHOD

Subjects

A relatively homogeneous group of eight high-status male fraternities at a large eastern university agreed to participate in the experiment which was represented as a study of the relationship between popularity and mode of functioning in a group problem-solving situation. The task was further presented as an interfraternity competition with publicity being given to the fraternity obtaining the highest composite score. As an added incentive, a prize of $25 was offered to the winning fraternity. At the termination of a weekly fraternity meeting the members ranked their fellow fraternity brothers in accordance with the following instructions:

> The following list contains the names of the members of your fraternity. Please rank these names in order of preference as a fraternity member. By preference is meant the extent to which you like the member, admire him, and are pleased to have him as a member of your fraternity. The one most preferred should receive the rank of 1, the second most preferred should receive the rank of 2, and so on. Continue your rankings until you have reached the ______ rank.[2] Please indicate your rank in the space provided next to each name.

In addition to the preference rankings, the men indicated on a 7-point rating scale the importance they placed on fraternity membership. Only those subjects whose ratings were in the upper half of the scale were included in the study. This latter procedure was employed in order to eliminate those subjects who placed a low valuation on their group membership and to keep constant the valuation of group membership of both high- and low-status subjects.

The four most-preferred and the four least-preferred men in each of eight fraternities comprised the experimental sample. The most-preferred members constitute the high-status group and the least-preferred members constitute the low-status group.[3]

---

[2] The number entered here was based on 80% of the total number of men in the particular fraternity.

[3] Fraternity officers consistently appeared among the most-preferred members, thus providing evidence for the validity of the preference rankings.

The mean valuation of membership ratings was almost identical for the high- and low-status groups, being 6.00 and 6.06, respectively. From each fraternity, subjects from each status group were then randomly assigned to one of the following four experimental treatment groups: high experimentally induced status-public knowledge, low experimentally induced status-public knowledge, high experimentally induced status-private knowledge, low experimentally induced status-private knowledge.

In addition, 15 subjects, randomly selected from the same high-preference–low-preference population as the 64 experimental subjects, were assigned to a control treatment.

Procedure

There were 16 experimental sessions, four men participating in each session. The four men in any one session, two high status and two low status, came from the same fraternity. Each fraternity was represented by two such groups, one of which was assigned to a public and the other to a private treatment condition. When the four members were assembled, they were informed that the study was concerned with how a group of friends operated in solving a problem together, and that although they would be separated in individual cubicles, they would be able to communicate with each other by typed anonymous messages at certain specific times, and their scores would be combined to yield a group total. At this point they were reminded of the prize and the publicity to be given to the winning fraternity. Those subjects assigned to the public treatment condition were told that after completion of the problem the group would be informed of the performance of each member, while subjects assigned to the private condition were told that their individual judgments would be kept confidential.

Upon entering the cubicle, an experimental assistant handed each man a folder presumably containing his answer sheet. Instead, included in a very prominent place in the folder was a fictitious sociometric summary rating sheet containing the ranks which the subject ostensibly received from his fraternity brothers. These fictitious ranks were used to induce conditions of *high experimentally induced status* and *low experimentally induced status*. Subjects assigned to the high experimentally induced status treatment found that their mean rank was in the upper decile of their group, while the low experimentally induced status subjects found that their mean rank was in the lower decile. In addition, the descriptive statements, "very popular, within the top 10," and "very unpopular, within the bottom 10," accompanied the corresponding numerical ranks.

In order for the subjects to have an opportunity to inspect the folder, the assistant left the cubicle to distribute the folders to the other subjects. To insure that the subject would find the fictitious ranking sheet, the assistant, before leaving the room, instructed him to fill in his name and fraternity on the answer sheet in the folder. After approximately 4 minutes, the experimenter entered the subject's cubicle with the experimental stimuli and proceeded with Trial 1. In almost every case the subject spontaneously indicated that he had not been given an answer sheet. At this point the experimenter inspected the folder, upon noting the ranking sheet feigned surprise and dismay, and apologized to the subject for this "inadvertent and careless" act of the assistant.[4]

Experimental Task

The experimental task, except for a few minor variations, was essentially the same as that used by Kelley and Shapiro (1954).

The task comprised a series of 11 problems, each consisting of the presentation of a pair of squares labeled "A" and "B." Each square initially contained approximately 100 dots, and the problem was to judge which square contained the larger number, there being too many dots to count in the allotted time. The subjects were informed that the correct answer would be the same throughout the series, either A or B being the right answer for all the problems. While initially it was not apparent which square had more dots (there being no difference, in fact, on the first trial), as the series continued, one square consistently lost a few dots on each presentation while the other square consistently gained a few.

After their judgments on the first problem were collected, the subjects were informed that they could only communicate with their groups by means of typed messages. The messages were intercepted and fictitious communications substituted. After the first and second judgment, the subject received two messages, each message stating that the smaller square (the incorrect one) was the correct answer. After the fourth judgment the subject received additional messages, one from each of his co-workers. These last messages confidently stated that the really incorrect square was the correct solution and urged the members of the group to agree upon this solution.

Since a subject's subsequent judgments may be influenced by persistence in his initial choice, the following counterbalanced design was employed. Half of the subjects were randomly assigned to a treatment in which their initial choice was made to be correct but opposed to the group (the *initially correct*, IC, condition), while the remaining half were randomly assigned to a treatment in which their initial choice was made to be incorrect and supported by the group (the *initially incorrect*, II, condition). In order to effect these conditions, alternate sets of stimuli and messages were prepared for each subject so that it was immaterial whether the subject selected Square A or Square B on the first trial. It should be noted that in all cases the group norm, as reflected in the fictitious messages, was

---

These officers were eliminated from the study in order to make the experimental manipulation of status more reasonable.

[4] During the postexperimental interviews, all subjects indicated that they had seen and were aware of their "fictitious" rank.

TABLE 1

PERCENTAGE OF INCORRECT RESPONSES ON EACH TRIAL FOR THE CONTROL AND EXPERIMENTAL GROUPS

Group	Trials									
	2	3	4	5	6	7	8	9	10	11
Control ($N = 15$)	60	53	33	13	20	20	20	0	0	7
Combined experimental ($N = 64$)	72	62	53	64	58	53	55	44	39	34
	χ^2									
	0.5	0.4	1.9	12.6***	6.9***	5.3**	5.9**	8.7***	6.9***	3.3*

* $p < .10$.
** $p < .05$.
*** $p < .01$.

incorrect. This control procedure constitutes a fourth variable, initial choice, which results in the formation of 16 experimental groups.

At the termination of the experimental session, the purpose of the experiment was explained in detail to the subjects, and they were assured that the rankings and the messages were fictitious. There was a good deal of interest and relief generated in these postexperimental sessions. Although the subjects were asked not to divulge or reveal the details of the experiment, as a precaution against contamination the two groups from the same fraternity were successively seen during the same evening. After the completion of the study, the "prize" money was awarded to the fraternity with the highest nonconformity (correct) score.

Control Group

The 15 subjects constituting the control group were individually given the series of experimental stimuli. These men did not work in a group context and did not participate in the competition for the prize. They received no messages, and they were not subjected to the experimentally induced status and public-private knowledge treatments. However, in all other respects, including the control for initial choice, the procedure employed for the experimental group was used for the control subjects.

Measure of Conformity, Nonconformity

A measure of conformity was obtained by summing for each subject the number of incorrect judgments for trials 2–11. Since the group pressure was always in the direction of favoring incorrect responses, a preponderance of incorrect responses in the experimental group over the control group may be attributed to conformity to this pressure.

RESULTS

Control versus Experimental Groups

A comparison between the control and combined experimental groups was made of the percentage of individuals in each group correctly responding on trials 2–11. The results of this comparison are presented in Table 1. The data indicate that on every trial the experimental group had a higher proportion of incorrect responses than did the control group. The difference between these groups became statistically significant at the fifth trial and maintained its significance until the eleventh trial. Even on this last trial where the contrast between the two stimuli was greatest, 34% of the experimental group still responded incorrectly. The difference between the control and experimental groups on this last trial fails to meet the .05 level because one control subject responded incorrectly on this trial. Thus, while the control data indicated that the stimuli were readily discriminable in the later trials, a significant proportion of the experimental subjects conformed to the fallacious group norm.

Effects of Experimental Variations

The mean conformity scores for each of the 16 experimental treatment groups are presented in Table 2, and the results of an analysis of variance of these conformity scores are presented in Table 3. These data indicate that the effects of the status and initial choice variables are significant at the .01 level. While the main effect of the public-private variable is not significant, it does yield a significant interaction with the status variable and also a significant triple interaction with status and initial choice. The p value for the main effect of the experi-

TABLE 2

MEAN NUMBER OF CONFORMING RESPONSES FOR EACH EXPERIMENTAL CONDITION

Group	High experimentally induced status				Low experimentally induced status			
	Public		Private		Public		Private	
	IC	II	IC	II	IC	II	IC	II
High status	2.8	7.8	4.3	2.8	1.5	7.8	0.8	3.0
Low status	9.5	7.5	5.5	8.8	2.3	6.0	6.5	9.0

Note.—$N = 16$ for each experimental condition.

mentally induced status variable falls between the .10 and .05 levels of significance. However, this effect is opposite to the predicted direction, the high experimentally induced status subjects showing more conformity than the low experimentally induced status subjects.

Interaction of the Status and Public-Private Conditions

The triple interaction of the status, public-private, and initial choice variables is plotted in Figure 1. These data indicate that the public-private condition has a significant influence upon conformity responses of the high-status subjects only when their initial choice is incorrect ($t = 3.1$, $p < .01$). Moreover, greater nonconformity of high-status as compared to low-status subjects is obtained under private conditions for both IC ($t = 2.2$, $p < .05$) and II ($t = 3.8$, $p < .01$) choices, and also under public conditions where the initial choice is correct ($t = 2.4$, $p < .05$). Only when their initial choice is incorrect and when they anticipate that their fellow fraternity members will be informed of their judgments do high-status subjects make as many incorrect judgments as low-status subjects.

TABLE 3

ANALYSIS OF VARIANCE OF CONFORMITY SCORES

Source	*df*	*MSS*	*F*
Status (A)	1	150.4	14.8****
Experimental status (B)	1	36.3	3.6*
Public–Private (C)	1	5.4	.5
Initial choice (D)	1	95.4	9.4***
A × B	1	1.9	.2
A × C	1	45.3	4.4**
A × D	1	5.0	.5
B × C	1	16.0	1.6
B × D	1	24.7	2.4
C × D	1	10.3	1.0
A × B × C	1	36.0	3.5*
A × B × D	1	.1	
A × C × D	1	56.6	5.6**
B × C × D	1	4.0	.4
A × B × C × D	1	15.0	1.4
Within	48	10.2	
Total	63	502.3	

* $p < .10$.
** $p < .05$.
*** $p < .01$.
**** $p < .001$.

The strength of the status variable under private conditions is sharply demonstrated when one examines the difference in conformity between the high- and low-status subjects at each trial. The pertinent data are presented in Table 4. The differences on the later trials are quite striking, particularly since the stimuli used on these trials are readily discriminable. Even on the last trial more than half of the low-status subjects in contrast to none of the high-status subjects conformed to the incorrect group norm. Hypothesis I, then, which asserts that high-status subjects conform less than low-status subjects, holds under private conditions and, as already noted, under public conditions when the initial choice was correct.

Effects of Initial Choice

It is apparent from the main effects and from the triple interaction that the correctness or incorrectness of the subject's first response was an important factor to control. When their initial choice was correct, subjects deviated significantly more from group norms than when their initial choice was the same as that of the group. The tendency to persist in one's initial choice can be clearly

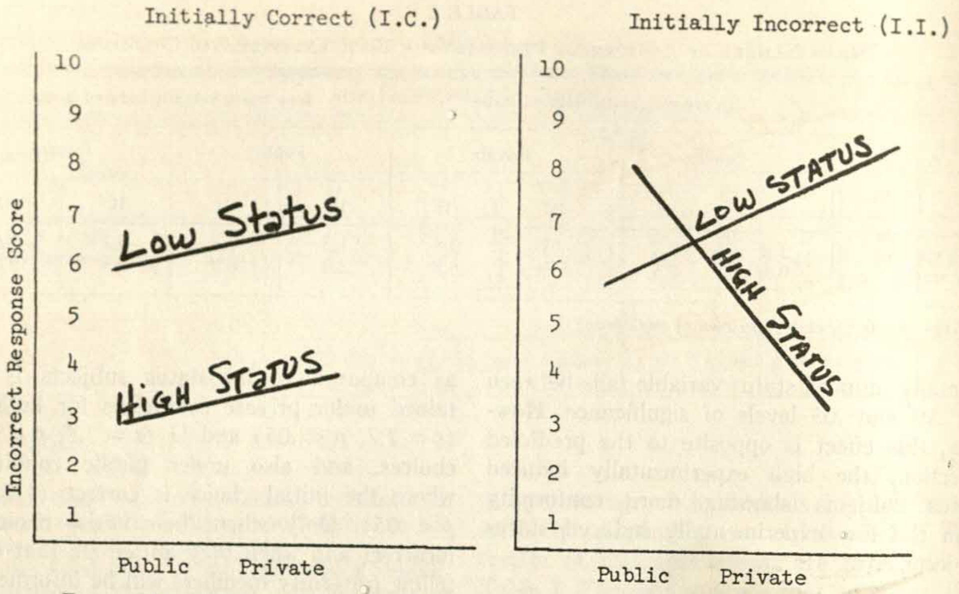


FIG. 1. The triple interaction of status, public-private, and initial choice variables.

observed in the control data which indicate that subjects whose initial choice was incorrect made significantly more errors on trials 2–11 than did the subjects whose initial choice was correct.

Effects of High and Low Experimentally Induced Status

It has already been indicated that the difference between the high and low experimentally induced status subjects is in a direction opposite to that predicted by the second hypothesis. A comparison of the proportions of high and low experimentally induced status individuals conforming on each trial revealed statistically significant differences ($p < .05$) between the high and low experimentally induced status subjects on trials 2–4. These data indicate that while on the first three trials the high experimentally induced status subjects conformed significantly more than low experimentally induced status subjects, as the trials progress the experimentally induced status variable loses its effectiveness. Some further information regarding the influence of this variable is provided by inspection of the triple interaction of the experimentally induced status,

TABLE 4

PERCENTAGE OF INDIVIDUALS CONFORMING ON EACH TRIAL IN THE HIGH-STATUS–PRIVATE AND LOW-STATUS–PRIVATE SUBGROUPS

Group	Trials									
	2	3	4	5	6	7	8	9	10	11
High status private ($N = 32$)	56	50	31	37	31	25	25	6	6	0
Low status private ($N = 32$)	94	81	63	81	81	75	87	69	63	56
	χ^2									
	4.6**	3.5*	3.1*	6.4**	8.1*	8.0*	12.7****	13.3****	11.2****	10.6***

* $p < .10$.
** $p < .05$.
*** $p < .01$.
**** $p < .001$.

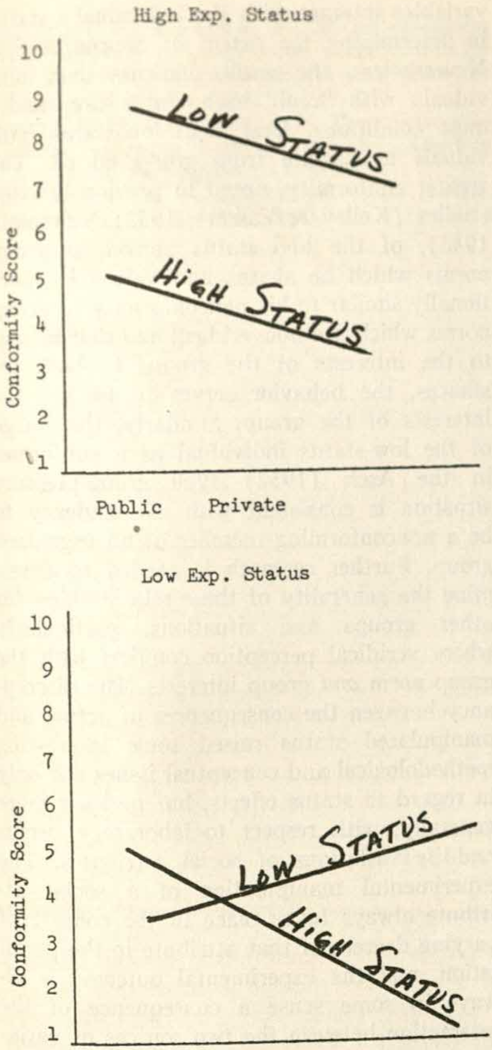


FIG. 2. The triple interaction of status, public-private, and experimentally induced status variables.

actual status, and public-private variables, which is graphically presented in Figure 2. It is apparent from Figure 2 that the marked effect of the experimentally induced status variable upon the low-status group under public conditions is primarily responsible for this triple interaction. Low-status subjects, when subject to the low experimentally induced status treatment, reacted in a nonconforming manner when they anticipated that their judgments would be made public to the members of their fraternity group. Low-status subjects under all other conditions showed significantly more conformity than under the low experimentally induced status-public condition. It is as if the low-status, low experimentally induced status subjects welcomed the opportunity under public conditions to reject the group norms.

DISCUSSION

One of the most striking features of the data is the great amount of conformity that was obtained in the experimental groups. Even on the very last trial, when the difference between the two stimuli was most sharply delineated, 22 out of the 64 experimental subjects adhered to the incorrect group norm.

From the postexperimental discussions and from the messages the subjects sent it was evident that an intense conflict existed for most subjects regardless of whether or not they conformed. The resolution of this conflict was in accord with the prediction made for the status variable. Except for the condition when their initial choice was incorrect and they expected public access to their judgments, high-status subjects conformed significantly less than low-status subjects. However, an opposite trend was obtained when status was experimentally induced. Both the quantitative results and the qualitative data indicate that the operations designed to induce high and low status must be evaluated in terms of their impact upon really high- and really low-status subjects.

When a high-status subject is confronted with a high ranking, his expectations and beliefs about his position in the group are confirmed. However, the high-status subject encountering a low ranking is confronted with a discrepancy between his self-perception and this new information. He is motivated to reconcile this discrepancy, and one way of doing so is to suspect the experimental procedure. Several subjects in this group were the only ones who gave voice to some suspicion of the procedure in their messages.

In contrast, the low-status subject discovering a fictitiously high ranking is less motivated to question it even though he may be surprised by his newfound position in the group. These subjects acted almost uniformly euphoric—whistling, laughing to themselves, and behaving in a very ingratiating, cooperative manner. The low-status individual, subjected to the low experimentally induced status treatment, however, finds his worst expectations unhappily confirmed. The reaction to this situation appears dependent upon whether these subjects are under private or public conditions. Under public conditions, their conformity scores show a significant decrease. One might interpret this decrease in conformity under public conditions as an assertion of their independence and perhaps rejection of their group.

The effects of the public-private variable are equally complex. The third hypothesis which predicted that private conditions would result in greater deviation from group norms is confirmed only for high-status subjects who were initially incorrect in their judgments, and, as noted above, the relationship is even reversed for the low-status, low experimentally induced status group. The inconsistent effect of the public-private variable observed in this study is similar to some of the findings reported by Kelley and Volkart (1952). The varied effect in the present study may in part be due to the familiarity of the subjects within each group who may have expected to discuss the experimental situation, regardless of whether they were under the public or private condition.

Throughout this discussion it has been apparent that the initial choice variable has been an important condition modifying the effects of the principal variables. In addition to the tendency to persist in one's initial choice, the relationship between the group norm and initial choice must also be taken into account. When the subject's initial judgment deviates from that of the group, there is an immediate recognition of a discrepancy between his judgment and that of the group, resulting in greater nonconformity. However, when the subject's initial judgment is the same as the group norm, the resulting mutual reinforcement tends to delay recognition of the conflict between the group norm and the properties of the stimuli.

It is evident that a number of situational variables interact with the individual's status in determining the extent of nonconformity. Nevertheless, the results indicate that individuals with "real" high status are, under most conditions, freer than low-status individuals to deviate from group norms. The greater conformity, noted in previously cited studies (Kelley & Volkart, 1952; Newcomb, 1943), of the high-status person to group norms which he shares and values is functionally similar to his nonconformity to group norms which are nonveridical and detrimental to the interests of the group. In both instances, the behavior serves to further the interests of the group. Similarly, the image of the low-status individual as a conformer in the Asch (1952) type group-pressure situation is consistent with his tendency to be a nonconforming member of an organized group. Further research is needed to determine the generality of these relationships for other groups and situations, particularly where veridical perception conflicts with the group norm *and* group interests. The discrepancy between the consequences of actual and manipulated status raised some interesting methodological and conceptual issues not only in regard to status effects, but perhaps more generally, with respect to laboratory versus real-life variations of social attributes. The experimental manipulation of a social attribute always takes place in the context of varying degrees of that attribute in the population, and the experimental outcome is always in some sense a consequence of the interaction between the two sources of variation even though the stable attribute may not be assessed.

REFERENCES

ASCH, S. E. *Social psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1952.

BERKOWITZ, L., & MACAULAY, J. R. Some effects of differences in status stability. *Human Relations*, 1961, **14**, 135–148.

FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups*. New York: Harper, 1950.

HOLLANDER, E. P. Conformity, status, and idiosyncrasy credit. *Psychological Review*, 1958, **65**, 117–127.

HOLLANDER, E. P. Competence and conformity in the acceptance of influence. *Journal of Abnormal and Social Psychology*, 1960, **61**, 365–369.

KELLEY, H. H., & SHAPIRO, M. M. An experiment on conformity to group norms where conformity is detrimental to group achievement. *American Sociological Review*, 1954, **19**, 667–677.

KELLEY, H. H., & VOLKART, E. H. The resistance to change of group anchored attitudes. *American Sociological Review*, 1952, **17**, 453–465.

NEWCOMB, T. M. *Personality and social changes*. New York: Dryden Press, 1943.

SABATH, G. The effects of disruptive and individual status on person perception and group attraction. *Journal of Social Psychology*, 1964, **64**, 119–130.

SCHACHTER, S. Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, 1951, **46**, 190–207.

(Received August 8, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 64-70

ATTITUDINAL EFFECTS OF COMMUNICATOR ATTRACTIVENESS WHEN ONE CHOOSES TO LISTEN[1]

RUSSELL A. JONES AND JACK W. BREHM

Duke University

The basic hypothesis of the present study concerned an interaction between communicator characteristics and whether or not Ss had volunteered to listen to the communicator. 2 predictions were based on this hypothesis: (a) Ss who had volunteered to listen to the persuasive communication would be influenced more by a negative than by a positive communicator; and (b) Ss who were "accidentally exposed" to the persuasive communication would be influenced more by the positive than by the negative communicator. Both predictions were borne out: the interaction was significant and analysis of the simple effects indicated that the positive and negative communicators differed significantly, in the appropriate directions, depending on whether Ss had volunteered to listen or were "accidentally exposed" to the communication.

If, in the absence of other forces, it may be assumed that a person would avoid listening to a persuasive communication which disagrees with his own point of view, then it should be true that if he is induced to choose to listen to the persuasive communication he will experience dissonance. According to dissonance theory (Festinger, 1957), the magnitude of dissonance experienced in regard to a given cognitive element is a direct function of the proportion of relevant cognitions which is dissonant. Thus, the magnitude of dissonance experienced from choosing to listen to a persuasive communication would be a function of (*a*) the content of the communication, (*b*) the magnitude of the force inducing the person to listen to the communication, and (*c*) non-content aspects of the communication situation which would be consonant or dissonant with having chosen to listen. With the content of the communication and the initial inducing force held constant, the magnitude of dissonance should vary inversely with, among other things, the personal attractiveness of the communicator. Given that choosing to listen to the communication constitutes a commitment, attitude change as an attempt to reduce dissonance should tend to be in the direction advocated by the communication. Specifically, then, one would expect that the more negative the personal characteristics of the communicator, the more attitude change there will be in the direction he advocates when one chooses to listen to him.

Powell (1965) has shown that attitude change is an inverse function of the credibility of a communicator who induces behavioral compliance in an unpleasant task. Similarly, Smith (1961) found that when people were induced to eat a disliked food, there was more positive attitude change toward it when the induction was by a "negative" than when it was by a "positive" experimenter. The same effect was subsequently shown more conclusively by Zimbardo, Weisenberg, Firestone, and Levy (1965). In all of these studies, however, the major source of dissonance was a commitment to engage in a distasteful overt behavior (giving blood or eating fried grasshoppers) rather than in simply listening to a persuasive communication. A basic point of the present study, then, was to see whether or not the behavior of choosing to listen would constitute a sufficiently strong commitment to provide dissonance effects in response to the communicator and communication.

---

[1] This study was supported in part in the form of a fellowship to the senior author under the National Defense Education Act, and in part by National Institute of Mental Health Grant MH-1122801 under the direction of the junior author. The authors wish to thank Joseph Mancusi, John Sensenig, Lloyd Stires, and Robert Wicklund for help in the preparation of experimental materials. We also wish to acknowledge the helpful suggestions regarding the procedure and this manuscript made by Darwyn Linder.

The present formulation assumes that when there is no choice about whether or not to listen to the persuasive talk, a person's tendency to listen and be persuaded is a direct function of the attractiveness of the communicator. Hence, the base line for the dissonance effects of positive and negative communicators is the differential effect of positive and negative communicators under no-choice conditions. Two steps are therefore necessary to demonstrate the dissonance effect: (*a*) Under no choice conditions, it must be shown that a positive communicator produces more attitude change than does a negative communicator; and (*b*) when a person chooses to listen, a negative communicator produces more attitude change than does a positive communicator *relative to the effects under no-choice conditions.* In other words, if attitude change is a direct function of communicator attractiveness under no-choice conditions, and if the dissonance hypothesis is correct, then there should be an interaction between the variables of choice and communicator attractiveness.

METHOD

College students listened to a recording which argued that intercollegiate sports are not an important aspect of college life under four experimental conditions: positive high choice, negative high choice, positive no choice, and negative no choice. In the positive conditions, the speaker on the tape indicated that he was very favorably impressed with today's college students, while in the negative conditions the speaker indicated just the opposite. In the high-choice conditions, subjects had volunteered to listen to the tape, in the no-choice conditions, subjects were seemingly accidentally exposed to the tape. A measure of agreement with the communicator's position, ratings of the communicator's characteristics, and a measure of the retained content of the persuasive communication were obtained after the subjects had listened to the tape.

The subjects were 60 female students in an introductory psychology course at Duke University. All subjects had signed up for "Filling Out Opinion Questionnaires." Students had to participate in a certain number of experiments as part of their course requirement, but were completely free to choose which experiments they signed up for. All subjects were run individually.

Pretesting a similar group of subjects the previous semester had established that nearly all agreed that "intercollegiate sports are an important aspect of college life." On the basis of this pretest, a persuasive communication was prepared which argued for the abolition of intercollegiate sports and quoted various statistics on the cost of intercollegiate sports to schools, declining attendance, and student apathy. Two tapes of the persuasive communication were prepared. On both tapes the credibility of the communicator was established by having him introduce himself as the head of the public relations department of the National Collegiate Athletic Association and make a few comments which demonstrated that he had been closely associated with intercollegiate athletics for a long time. The contents of the persuasive communication proper were given as the findings of a recent study he had conducted for the NCAA. The two tapes of the persuasive communication were used to establish the differential characteristics of the communicator (his positiveness or negativeness). On both tapes, after the communicator had established his credibility he noted that for the study he had just completed for the NCAA, undergraduates were used as contacts at each school surveyed. The difference between the positive and negative tapes was introduced by the communicator saying (on the positive tape):

> I was a little apprehensive at first about using undergraduates, but as the study got started I found that undergraduates were extremely competent and quite concerned, in general, with doing a good job. As a result of this experience I've come to the conclusion that we too often underestimate both the ability and maturity of today's college student. It was really quite a rewarding experience for me.
>
> As for the study itself . . . .

On the negative tape the communicator said:

> One thing I did learn from this study was to never use undergraduates again. I found that undergraduates were very difficult to work with or to get any work out of. They made so many stupid mistakes that it was impossible to guide them in any meaningful sense. Using them resulted in a waste of time and money and I ended up doing three times as much work as I had anticipated because it got to the point that I couldn't depend on them for accurate reports.
>
> As for the study itself . . . .

The remainders of the tapes were identical, and the communicator took a strong stand for doing away with the type of intercollegiate sports now so prevalent on college campuses."

Because the procedures used differed rather radically for the high-choice conditions and no-choice conditions, the four experimental conditions mentioned above might be more appropriately categorized as two different, but comparable, experiments of two conditions each (a positive communicator condition and a negative). Because the no-choice conditions were deemed more likely to be affected by suspicion, these two conditions were run first. Half of the subjects in a control condition (who heard no tape) were interspersed with the subjects in the low-choice conditions and half were interspersed with the (later-run) subjects in the high-choice conditions. The entire experiment was completed within 11 days.

The experiment required the use of three rooms: a waiting room, a room in which the subject remained during the experiment, and a room in which a tape recorder was set up—directly across the hall from the experimental room.

No-Choice Conditions

In the no-choice conditions, subjects were run in the following manner. When a subject reported to the waiting room—the outer office in a suite of small rooms used for experiments—the secretary called the experimenter over an intercom and informed him that a subject had arrived. The experimenter then started the recorder with either the positive or negative version of the persuasive communication. Both versions of the tape had 4 minutes of "filler," dealing with sociological analysis, before the persuasive communication began. During the "filler" the experimenter left the room in which the recorder was playing—leaving the door ajar and the volume up enough so that the recording could be heard in the experimental room when the door to the experimental room was open—went to the outer office, accompanied the subject into the experimental room, and closed the door. The experimenter then checked the subject's name and engaged the subject in conversation for a minute or two. By checking a stopwatch, the experimenter was able to determine how long before the "filler" on the tape ran out. Approximately 30 seconds before the persuasive communication was to begin the experimenter told the subject that "we might as well let you get started on these questionnaires." He then picked up a questionnaire from a stack, and as he was handing it to the subject he "discovered" that the secretary had left him the wrong opinion questionnaires. He apologized to the subject for the delay but said that he would have to go get the right forms and left the experimental room—leaving the door open so that the subjects could hear the persuasive communication being played on the recorder in the room across the hall.

The persuasive communication itself was 6½ minutes long. In order to avoid the experimenter's absence being suspiciously long, when the communication had been playing 4 minutes the experimenter returned to the experimental room with a stack of forms, laid them on a table, picked up the other stack (the "wrong" forms), mumbled to himself that he had "better put these back in case they need them," and left again with the door still open. After the persuasive communication had been playing for 6 minutes and 10 seconds, the experimenter returned to the experimental room, closed the door, and again apologized to the subject for the delay. The experimenter then picked up a copy of the opinion questionnaire and turned to the subject with the following remarks:

> You'll notice that on this first questionnaire we don't ask for your name—and we don't need your name. I do want to reassure you that no one will see how you answer the particular items. After everyone is through we prepare a group summary for each item and we don't keep a record of who filled out which questionnaire. We would like that information at the top (sex, class, major), if you don't mind. Just let me know when you get to the end of that one and I'll give you the second one.

The opinion questionnaire consisted of 32 statements each accompanied by a seven-division scale with the end divisions labeled "strongly disagree" and "strongly agree." Below each item was a five-division confidence scale on which the subjects could indicate how confident they were of their position on that item. All 32 items were "school related." The critical item concerning intercollegiate athletics was imbedded in the questionnaire. As soon as subjects had finished the questionnaire they were given a second questionnaire with the following comments:

> Now about this second questionnaire, I have a little confession to make. We really aren't so interested in opinions as we are in certain aspects of the attention and memory processes. If you will just fill out this next questionnaire I'll be glad to answer any questions you might have as soon as you're through.

The second questionnaire contained five multiple-choice questions pertaining to information contained in the persuasive communication, one question and scale concerning the amount of choice the subject perceived in being able to refuse to listen to the persuasive communication, and 10 polar-adjective scales on which the subject rated the communicator.

When the subjects had completed the second questionnaire the experimenter probed for suspicion and then explained the experiment and the deceptions. All subjects were asked not to talk about the experiment for "a month or so."

High-Choice Conditions

When a subject who had been assigned to the high-choice conditions reported to the waiting room, the experimenter came out—without starting one of the tapes—and accompanied her to the experimental room. He then engaged the subject in conversation for a few minutes before going into the instructions for the high-choice conditions—which were of the forced-compliance variety.

In the instructions, subjects were told that recent research had indicated that if people listened to an opinionated statement before filling out opinion questionnaires they tended to give "more stable" responses and to take the task more seriously. They were told that the present experiment had been designed to check on the "odd finding that it doesn't seem to matter what the opinionated statement is about, you still seem to get the same effect."

The experimenter informed them that a large number of tapes on different issues had been collected—but not labeled—and that one was played for each subject.

It was repeatedly emphasized to the subject that some students had found some of the tapes objectionable and that the subject did not have to listen if she didn't want to. It was also emphasized that the experimenter did not know which tape was to be played and, hence, could not guarantee whether or not it would be objectionable to the subject. After reiterating that the subject would not be upsetting anything if she "would just rather not listen," the experimenter paused and waited for some verbal indication from the subject as to whether or not she wanted to listen. He then repeated again that the subject did not have to listen if she didn't want to and waited for a second indication from the subject.

After the second affirmative answer, the experimenter explained that he had set up the recorder in the room across the hall but that he would open the door and she would be able to hear the tape where she was seated. He also said that he had to get the "right" forms (as in the no-choice conditions), but would be back before the tape was over. The experimenter then went out, leaving the door open, went into the room across the hall, started the tape (without the "filler"), and walked off down the hall as in the no-choice conditions. From this point, the procedure was identical with that of the no-choice conditions (above).

Control Condition

Subjects in the control condition simply came to the experimental room, filled out the opinion questionnaire, were told about the experiment, and sworn to secrecy. Half (6) of the control subjects were run during the time that the no-choice conditions were being run and half during the time that the high choice conditions were being run.

RESULTS

Three subjects were eliminated from the analysis: one was not an undergraduate, one was aware of the deceptions, and one misperceived the choice manipulations.

Checks on Manipulations

Evidence from an item on the recall questionnaire indicates that the choice manipulation was effective. The relevant item asked subjects to check on a 7-point scale what opportunity they had had to refuse to listen to the tape (the discrepant communication). With "no opportunity" scored as 1, and "several" scored as 7, no subject in the no-choice conditions checked above 3 on the scale and only one subject in the high-choice conditions checked below 7. There was no overlap between the two conditions and the difference between them is highly reliable. The attractiveness of the communicator had no effect on the degree of perceived choice.

To check on the attractiveness of the communicator, subjects were asked to rate the speaker on 10 polar-adjective scales. The negative end of each scale was scored as 1 and the positive end as 7. A summation score was obtained for each subject on all 10 scales and an analysis of variance carried out. This analysis indicates that the attractiveness manipulation reliably affected the evaluations of the speaker ($F = 15.16$, $p < .001$), and its effectiveness was not a function of the choice manipulation. In both this analysis and the analysis of the dependent variable the assumptions of the analysis of variance were violated[6] to some extent because assignment of subjects to choice conditions was not random.

In order to inspect the differences in the perception of the speaker more closely, data on each of the 10 scale items are presented in Table 1. All dimensions in Table 1 except Intelligent–Unintelligent and Well read–Poorly read yield significant differences between groups exposed to the positive and negative speakers at the .005 level or better. The few remarks made by the speaker about

TABLE 1

ITEMS RATING SPEAKER

Polar adjectives	Positive groups		t	Negative groups	
	$\bar{X}$	s^2		$\bar{X}$	s^2
Competent–Incompetent	5.375	2.33	2.37	4.250	3.06
Warm–Cold	5.167	1.88	3.80	3.625	2.51
Intelligent–Unintelligent	5.292	1.87	1.09	4.833	2.41
Approachable–Unapproachable	5.500	3.30	2.86	4.042	2.91
Interesting–Dull	6.000	.61	3.74	4.542	3.04
Generous–Stingy	5.167	1.97	4.92	3.250	1.67
Friendly–Distant	5.500	2.17	2.45	4.417	2.51
Good sense of humor–Poor sense of humor	5.500	1.48	3.52	4.083	2.43
Well read–Poorly read	4.917	2.51	1.11	4.417	2.34
Modest–Arrogant	4.208	2.78	3.08	2.792	2.34

Note.—$N = 24$, both groups.

undergraduates clearly yielded differing perceptions of the speaker on the part of the subjects.

Because of the nature of the no-choice conditions, some check on whether or not subjects in these conditions paid sufficient attention to the discrepant communication was deemed necessary. Pretesting had revealed that when subjects in the negative no-choice condition heard the negative remarks about undergraduates, some tended not to listen to the main body of the communication because they felt it "wasn't meant for me to hear." To check on whether or not subjects in the different conditions did differentially attend to the communication, five items on the content of the persuasive communication were included on the recall questionnaire. The data in Table 2 show that there were no differences even approaching significance in the number of correct answers to the questions on the content of the persuasive communication. These data tend to rule out any alternative explanations in terms of differential attention to the persuasive communication. It should also be noted that the recall questions seem to have been sufficiently difficult to have detected any differences in attention since in no condition did the scores on the recall items approach the maximum possible.

Tests of Hypotheses

The two main hypotheses being investigated concerned the differential reactions to communications by positive and negative communicators, depending upon whether one perceives a high or low degree of choice in determining whether or not one listens to the communication. The main dependent measure was an item on the opinion questionnaire which read, "Intercollegiate sports are an important aspect of college life." The item was accompanied by a 7-position scale with "strongly disagree" scored as 1 and "strongly agree" as 7. Hence, the lower the score on the item the more the agreement with the communication, since the communication argued that intercollegiate sports are not an important aspect of college life. Table 3 presents means and variances of the attitude scores for the five conditions.

TABLE 2

MEAN RIGHT ANSWERS IN THE FOUR CONDITIONS ON FIVE RECALL QUESTIONS

Condition	Communicator	
	Positive	Negative
High choice	3.33[a]	2.83
No choice	2.83	2.91

Note.—$N = 12$, all conditions.
[a] Perfect recall = 5.

TABLE 3

MEAN AGREEMENT WITH THE PERSUASIVE COMMUNICATION

Condition	$\bar{X}$	s^2
Control	6.00	1.091
No choice		
Negative	6.25	1.750
Positive	4.75	2.386
High choice		
Negative	5.25	2.023
Positive	6.58	.265

Note.—The lower the mean the more the agreement with the communication.

As can be seen in Table 3, subjects in the negative high-choice and positive no-choice conditions come to agree with the speaker more than subjects in any of the other conditions. However, the positive low-choice condition is the only one of the four experimental conditions which is significantly different from the control condition ($t = 2.35$, $p < .02$). The difference between the control and the negative high-choice condition approaches significance, but is not reliable ($t = 1.47$, $p > .10$).

An analysis of variance on the attitude scores for the four experimental conditions demonstrated that the predicted interaction between speaker characteristics and choice condition is highly significant ($F = 15.66$, $p < .001$). Evidence from the analysis of simple effects (Winer, 1962) provides even stronger support for the two main hypotheses. In the high-choice condition, there is more agreement with the negative communicator than with the positive ($F = 6.937$, $p < .05$), and in the no-choice condition, there is more agreement with the positive communicator than with the negative ($F = 8.778$, $p < .01$).

DISCUSSION

Both hypotheses were supported. Specifically, having volunteered to listen to a persuasive communication, subjects agreed more with a negative than a positive communicator and when "accidentally exposed" to a discrepant communication, subjects agreed more with a positive than a negative communicator. These results both support and extend the findings of Zimbardo et al. (1965) and Smith (1961), in that in the present study the behavior of choosing to listen did constitute a sufficiently strong commitment to provide dissonance effects, whereas in the other studies the major source of dissonance was a commitment to engage in a distasteful overt behavior. A basic assumption underlying the present study is that in the absence of other forces a person would avoid listening to a persuasive communication which disagrees with his own point of view. This assumption was necessary in order to specify that dissonance would be aroused when a person is induced to listen to a discrepant communication. It should be noted that, as stated above, this assumption is not refuted by evidence that in certain postdecisional "choice among alternatives" situations, subjects do not necessarily avoid dissonance-increasing information. Freedman and Sears (1965), in a recent review of selective exposure, were quite unimpressed with the evidence for the selective exposure hypothesis. But even if the present study were taken as being of the selective exposure variety—which it is not—Freedman and Sears pointed out that subjects prefer supportive information when they believe the information content is unfamiliar. Hence, being induced to listen to nonsupportive unfamiliar information would still arouse dissonance.

Another aspect of the present study which requires attention is the relation of the control group to the experimental groups. Although the control group does seem fairly comparable to the no-choice conditions, it is less clear that the control group really constitutes an adequate "control" for the high-choice conditions. A more appropriate control for the high-choice conditions would entail a procedure similar to that for the high-choice groups, but with the persuasive communication about some topic unrelated to intercollegiate sports. Even though such a control condition would have been more appropriate for the high-choice conditions, it is difficult to see how the high-choice procedure per se influenced attitudes toward intercollegiate athletics.

The obtained pattern of results in the present study could have occurred for other reasons; that is, what Orne (1962) called the demand characteristics of the experimental situation. It is possible that subjects "divined" what was expected of them. This seems rather unlikely, however, for several reasons. First, the subjects were rather experimentally naive. For nearly all the subjects this was the first social psychological experiment that they had been in. Second, in the no-choice conditions, spontaneous remarks from the subjects when the experimenter handed them the recall questionnaire on the persuasive communication indicated that the subjects were genuinely surprised and dismayed that they had been expected to listen to the "accidentally" overheard communication. While the problem of the cooperative subject seems not to be relevant to the control condition or to have influenced the no-choice conditions, there is still the possibility that subjects in the high-choice conditions agreed with the negative communicator more than the positive communicator simply to show the experimenter how open-minded and objective they were. One thing which argues against this possibility, although rather weakly, is the fact that the opinion questionnaires did not ask for the subject's name, and subjects were verbally reassured that the data would remain anonymous and "group summaries" would be prepared for each item. Also, it seems that if this pseudocooperation had been operating in the high-choice conditions it would have also been reflected on the recall questionnaire; that is, subjects would have rated the negative speaker as more negative and the positive speaker as more positive than did subjects in the no-choice conditions. Data from the rating scales indicate that this is not so; subjects in the high-choice conditions rated the communicator no differently than did subjects in the no-choice conditions. Hence, while the "demand characteristics of the experimental situation" could conceivably have produced the present results, it seems rather unlikely that they did.

Since there was some heterogeneity of variance on the experimental attitude scores and the present study used an after-only design, there is the possibility that the results are due to differences among the conditions in initial attitude. There is no way of definitely ruling out this possibility, but evidence from the opinion questionnaire administered to all subjects indicates that the groups assigned to the various conditions were homogeneous with respect to means and variances on other opinion scores.

REFERENCES

Festinger, L. *A theory of cognitive dissonance.* Stanford, Calif.: Stanford University Press, 1957.

Freedman, J. L., & Sears, D. O. Selective exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 2. New York: Academic Press, 1965.

Orne, M. On the social psychology of the psychological experiment with particular reference to demand characteristics and their implications. *American Psychologist,* 1962, 17, 776–783.

Powell, F. H. Source credibility and behavioral compliance as determinants of attitude change. *Journal of Personality and Social Psychology,* 1965, 2, 669–676.

Smith, E. E. The power of dissonance techniques to change attitudes. *Public Opinion Quarterly,* 1961, 25, 626–639.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Zimbardo, P. G., Weisenberg, M., Firestone, I., & Levy, B. Communicator effectiveness in producing public conformity and private attitude change. *Journal of Personality,* 1965, 33, 233–255.

(Received August 8, 1966)

Manuscripts Accepted for Publication in the *Journal of Personality and Social Psychology*

Seating Arrangements and Status: Dale F. Lott* and Robert Sommer: Department of Psychology, University of California, Davis, California 95616.

Decision Freedom as a Determinant of the Role of Incentive Magnitude in Attitude Change: Darwyn E. Linder, Joel Cooper, and Edward E. Jones*: Department of Psychology, Duke University, Durham, North Carolina 27706.

Bilingual Semantic Merging and an Aspect of Acculturation: Margaret J. Earle*: Department of Psychology, Taylor Building, University of Aberdeen, Aberdeen, Scotland, Great Britain.

Dimensions of Verbal Racial Attitudes: Their Identification and Measurement: John J. Woodmansee and Stuart W. Cook*: Institute of Behavioral Science, University of Colorado, Boulder, Colorado 80304.

Individual Differences and Perceptual Defense in the Absence of Response Bias: Richard R. Bootzin* and Mark W. Stephens: Department of Psychology, Purdue University, Lafayette, Indiana 47907.

Behavioral Receptivity to Dissonant Information: Timothy C. Brock* and Joe L. Balloun: Psychology Department, Ohio State University, 1945 North High Street, Columbus, Ohio 43210.

Determinant of Preference among Modes of Dissonance Reduction: Elaine Walster,* Ellen Berscheid, and Andrew M. Barclay: Laboratory for Research in Social Relations, University of Minnesota, Minneapolis, Minnesota 55455.

Belief Similarity versus Race as Determinants of Reactions to Negroes by Southern White Adolescents: A Further Test of Rokeach's Theory: Chester A. Insko* and James E. Robinson: Department of Psychology, University of North Carolina, Chapel Hill, North Carolina.

Social Experience and Social Schemata: Laura Weinstein*: Department of Psychology, George Peabody College for Teachers, Nashville, Tennessee 37203.

Dissonance Arousal and Task Performance in an Incidental Verbal Learning Paradigm: Michael S. Pallak, Timothy C. Brock, and Charles A. Kiesler*: Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510.

Recognition of Taboo Words as a Function of a priori Probability: Donald Daniel Dorfman*: Psychology Department, San Diego State College, San Diego 15, California.

Respondent's Own Attitude as a Source of Cognitive Bias in the Representation of Simple Social Structures: Eugene Burnstein*: Institute for Social Research, The University of Michigan, Ann Arbor, Michigan 48104.

Kinetic versus Social Schemas in Figure Grouping: David W. Lewit* and Virginia D. Joy: 4104 Institute for Social Research, The University of Michigan, Ann Arbor, Michigan 48104.

Compliance without Pressure: The Effect of Guilt: Jonathan L. Freedman,* Sue Ann Wallington, and Evelyn Bless: Department of Psychology, Stanford University, Stanford, California 94305.

Amount of Experience in Experiments as a Determinant of Performance in Later Experiments: David S. Holmes*: Department of Psychology, The University of Texas, Austin, Texas 78712.

Longitudinal Study of the Origins of Achievement Strivings: Shelia C. Feld*: Mental Health Study Center, National Institute of Mental Health, 2340 University Boulevard, Adelphi, Maryland.

---

* Asterisk indicates author for whom address is supplied.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 71–78

EFFECT OF VICARIOUS SCHEDULING ON THE VERBAL CONDITIONING OF HOSTILITY IN CHILDREN[1]

RAYMOND DITRICHS, SEYMORE SIMON, AND BARRY GREENE[2]

Northern Illinois University

2 experiments investigated the direct transfer effects of exposing Ss (7th, 8th, and 9th graders) to different vicarious schedules of hostile verb choices in a Taffel-type sentence-construction task. Results of Exp. I indicated that, with the exception of Ss who listened to a model construct progressively fewer reinforced hostile sentences, vicarious experience facilitated the subsequent conditionability of neutral verb choices. Relative to the other vicarious groups in Exps. I and II, Ss who listened to a taped acquisition sequence showed greater increases in the use of hostile verbs early in transfer and showed pronounced conditionability for hostile verb choices over all direct-reinforcement trials. In Exp. II, the conditioning of hostile verb choices was found to be more difficult when Ss were required to employ 1st-person pronouns rather than 3rd-person pronouns in their sentence constructions; however, conditioning in the 1st person was facilitated by prior vicarious experience. In both experiments few Ss were classified as being "aware," and no performance differences attributable to awareness were evident.

Recently verbal conditioning paradigms have been utilized to study vicarious learning in human subjects. Employing a modified Greenspoon (1950) procedure, Kanfer and Marston have examined the effects of such variables as the presence or absence of vicarious reinforcement, changes in the number and proportion of vicarious reinforcements, the addition of direct reinforcement, instructions to subjects, and reinforcement combinations on both the vicarious acquisition and extinction of selected verbal response classes (Kanfer & Marston, 1963; Marston, 1964, 1966; Marston & Kanfer, 1963). Other experimenter's have reported vicarious verbal acquisition utilizing Taffel-type sentence-construction tasks (Fadigan, 1961; Hurlock & McCain, 1964; Marlowe, Beecher, Cook, & Doob, 1964).

In all of the above studies, irrespective of the verbal conditioning paradigm employed, subjects either observed another subject (the model) undergo conditioning or listened to a simulated conditioning session during the vicarious acquisition. Thus, in all instances, subjects were exposed to progressive increases in the use of selected response classes by a model. In the present study, the distribution of the reinforced class of responses over the vicarious trials was varied, keeping their number and proportion constant, and the effect of this variation on subsequent conditioning was investigated. Children were employed as subjects in a Taffel-type verbal conditioning task with hostility as the discriminative dimension. While a number of studies have demonstrated that hostile or aggressive nonverbal responses can be learned vicariously in children (Bandura & Walters, 1963), direct evidence for the vicarious learning of hostile verbal responses is virtually nonexistent.

EXPERIMENT I

Method

Subjects. Subjects were 75 male and 75 female eighth- and ninth-grade students. Ten subjects, five of each sex, were randomly assigned to each of 15 experimental groups.

Stimulus materials. A pool of 650 past-tense verbs was first presented to 50 male and 50 female introductory psychology students who were asked to categorize each verb as being either "hostile" (H) or

---

[1] This research was supported by funds from United States Public Health Service Grant MH 08917. We are indebted to the staffs of the DeKalb, Sycamore, University, Illinois; and Clinton, Iowa, junior high schools for their assistance in providing subjects. Biruta Ditrichs and Nancy Jamison assisted in collecting and analyzing the data of Experiment II.

[2] Now at Monmouth College.

"neutral" (N). The 132 most frequently chosen H verbs and a similar number of highly rated N verbs, equated with the H verbs for Thorndike-Lorge frequency of occurrence on the G count, were selected from these ratings.

Following random pairings of these 132 H and 132 N verbs, 72 of the H-N pairs were randomly chosen, and each pair, together with the pronoun "they," was typed in primary letters on a 3 × 5-inch index card. The two verbs were placed to the right of the pronoun, one above the other. On half of the cards the H verb was in the first position, and on the other half of the cards the order was reversed.

The verbs of the remaining 60 H-N pairs were used in constructing sentences, 4–10 words in length. All sentences began with the pronoun "they," followed by either an H or N verb. Thirty of the sentences employed H verbs and 30 sentences used N verbs. The sentences were recorded on tape, with a male graduate student serving as the experimenter and a female junior high school student serving as the subject. On each recorded trial, the experimenter said the pronoun "they," followed by an H and an N verb, and the subject responded with the predetermined sentence. On a random half of the 60 taped trials, the experimenter presented the H verb first, and on the other half of the trials the N verb preceded the H verb. Instructions to the taped subject also were recorded.

Procedure. Subjects were run individually in a room provided by the school. Each subject was seated at a table facing the experimenter, the same male graduate student who had served as the experimenter in the recorded sessions.

In the instructions, the experimenter described a sample card and told the subject that his task for each card was to use the pronoun "they" and one of the two verbs on the card to make up a sentence. Twelve operant-level trials were then given, with all subjects treated alike in this initial phase of the experiment. On each trial, the experimenter exposed a card and the subject responded with a sentence. The same set of 12 cards was presented to all subjects; however, the order of the cards varied randomly from subject to subject.

Following the operant-level trials, four groups of subjects were told that some weeks ago a tape recording had been made of a student making up some sentences in the same way as they had just made up sentences. They were then asked to simply listen to the recorded sentences. For all groups, the experimenter remained in the room during the recorded playbacks.

The four groups differed from each other with respect to the tape which they received. For subjects in Group X-1, the first vicarious block of 12 trials consisted only of N sentences; that is, the subject on tape used only the N verbs to make up the sentences. On vicarious Blocks 2, 3, 4, and 5 the percentages of H sentences constructed were 25, 50, 75, and 100, respectively. All of the H sentences were verbally reinforced on tape by the experimenter saying "good." This condition thus constituted an ascending vicarious sequence. For subjects in Group X-2, a descending vicarious-sequence group, Blocks 1–5 contained 100, 75, 50, 25, and 0 percentages of H sentences, respectively. As with subjects in Group X-1, all sentences containing H verbs were verbally reinforced by the experimenter on tape. Subjects in Groups X-3 and X-4 were treated alike in that a random half of the sentences constructed in each taped block used H verbs and half used N verbs. However, for subjects in Group X-3 these sentences were verbally reinforced on tape, while for subjects in Group X-4 none of the sentences were reinforced. Thus, subjects in all of the X groups received an equal number of vicariously constructed sentences containing 50% H verbs and 50% N verbs. Furthermore, the number of vicarious reinforcements was equated for Groups X-1, X-2, and X-3, since all sentences containing H verbs were verbally reinforced on tape. These latter groups differed only in terms of the vicarious schedule that they received. A final group of subjects did not listen to any of the tape conditions and served as a control for vicarious experience per se in the experimental situation.

Immediately following the tape conditions, subjects in the X groups were again instructed to make up sentences from the materials on cards; control subjects received similar instructions immediately following operant-level trials. Sixty transfer trials were administered to all subjects. One-third of the subjects in each of the five groups were verbally reinforced for choosing the H verb, one-third of the subjects were reinforced for choosing the N verb, and the remaining one-third of the subjects received no reinforcement for their sentence constructions. Thus, for subjects in Groups X-1, X-2, and X-3, the three transfer conditions were, respectively, direct transfer, counterconditioning, and extinction. Reinforcement consisted of the experimenter saying "good." The same 60 cards, presented to each subject in a different random order, were used in the transfer phase. The experimental design was a 5 (Tape Conditions) × 3 (Transfer Conditions) factorial with independent groups of subjects in each of the cells.

The following postexperimental questionnaire was administered to all subjects in the X groups upon completion of the transfer phase:

1. How did you decide which of the two words to choose for your sentences?
2. Did anything that you heard on the tape influence your choice of words? (a) What? (b) How did it influence your choice?
3. Did you notice my saying anything on the tape besides the words which I presented to the girl to choose from? (a) Did you notice that I said "Good"? (b) When did I say "Good"?
4. Was there any other time that I said "Good"? (a) When? (b) Did it in any way influence the choices you made? (c) How?

The questionnaire administered to the control subjects was appropriately modified and contained no reference to listening to tape.

RESULTS

All experimental sessions were recorded on tape and subsequently transcribed. Prior to statistical analyses, two of the authors independently examined the postexperimental questionnaires and rated each subject as being either "aware" or "unaware" of the reinforcement contingency. A subject was rated as aware if he either verbalized the reinforcement contingency or stated that he was expected to use the hostile/neutral verbs in constructing the sentences. Results indicated that a large majority of subjects could verbalize the fact that the experimenter said "good" for the sentence constructions in the transfer phase, as well as for the sentences constructed by the model on tape. However, only two subjects, one subject in Group X-3 and one subject in Group C, were classified by both raters as being aware of the reinforcement contingency. No rating disagreements were obtained. Therefore, these questionnaire data were ignored in subsequent analyses.

To determine whether the randomly constituted groups were comparable prior to the introduction of the experimental treatments, a Tape Conditions × Transfer × Sex analysis of variance was conducted on the operant level scores. In this analysis, as well as in subsequent analyses, the criterion measure was the number of sentences constructed using hostile verbs per block of 12 trials. The main effect for sex was found to be the only statistically significant ratio (F (1, 120) = 4.86, $p < .05$). The means for the male and female subjects were 5.1 and 4.4, respectively.

To examine the immediate transfer effects of vicarious experience a Tape Conditions × Transfer × Sex analysis of covariance was conducted on Block 1 scores with adjustments made for operant level performance. Results indicated that the main effect for tape conditions was the only significant source of variance (F (4, 119) = 3.99, $p < .01$). The results of t tests on the adjusted condition means[3] revealed that subjects listening to an ascending vicarious sequence (Group X-1) chose significantly more H verbs on Block 1 than did subjects listening to either a descending sequence (Group X-2; $p < .001$), a randomly reinforced sequence (Group X-3; $p < .01$), or a randomly nonreinforced sequence (Group X-4; $p < .05$). Other two-group comparisons were not statistically reliable. The adjusted means for Groups X-1, X-2, X-3, X-4, and C were 6.7, 5.0, 5.5, 5.6, and 5.8, respectively.

A Tape Conditions × Transfer × Sex × Blocks analysis of variance on the transfer phase yielded significant main effects for tape conditions, F (4, 120) = 9.40, $p < .001$ and transfer, F (2, 120) = 17.70, $p < .001$. The following interactions were also found to be significant: Tape Conditions × Blocks, F (16, 480) = 1.86, $p < .05$, Transfer × Blocks, F (8, 480) = 2.23, $p < .05$, Sex × Blocks, F (4, 480) = 2.82, $p < .05$, and Tape Conditions × Transfer × Blocks, F (32, 480) = 1.50, $p < .05$. Tukey's T method indicated that Group X-1 differed significantly from each of the other four groups ($p < .01$); the other group differences were not statistically reliable. The group means were 6.7, 5.2, 5.8, 5.2, 5.7 for Tape Conditions X-1, X-2, X-3, and X-4, and C, respectively. A similar multiple-comparisons analysis indicated that subjects reinforced for selecting H verbs constructed more H sentences than did subjects receiving either reinforcement for N verbs or no reinforcement ($p < .01$ in both instances); the latter two groups also differed significantly from each other ($p < .05$). The means for the H-reinforced, N-reinforced, and no-reinforcement groups were, respectively, 6.4, 5.1, and 5.7.

Since sex differences from block to block were found to be small and nonsystematic, no further analyses of these effects were made. The significant Tape Conditions × Transfer × Blocks interaction is of primary interest in this experiment, however, since it indicates that performance over direct conditioning trials was dependent upon previous vicarious experience. Figure 1 presents this interaction. To evaluate the nature of this effect, Transfer × Blocks interactions were computed for each tape condition separately. Results indicated that the interaction was significant for Group X-1, $F\ (8, 96) = 3.87$, $p < .001$, and tests for trend showed significant linear trends over blocks for the subgroups reinforced for selecting H and N verbs (p's $< .001$, $.01$, respectively); no significant trend was obtained for the subgroup receiving no reinforcement in transfer. The Transfer × Blocks interactions were not statistically significant for Groups X-2 and X-3. For Group X-4, a significant Transfer × Blocks interaction was obtained, $F\ (8, 96) = 3.14$, $p < .01$, and a significant linear trend was found for the subgroup reinforced for choices of N verbs ($p < .001$); the H and no-reinforcement subgroups showed no reliable trends. No significant findings were obtained for subjects receiving no vicarious pretraining (Group C).

---

[3] Since the standard error of a difference between two adjusted means depends on the difference of the covariate means, Tukey's T method of multiple comparisons is not appropriate. Therefore t tests were run.

To determine whether the lengths of the sentences constructed by subjects in the transfer phase varied with the reinforced class of verbs, product-moment correlations were computed between mean sentence lengths and the number of reinforced responses on Block 5. The obtained values of r were .09 and .08 for the subjects reinforced for choices of H and N verbs, respectively. The correlations between Block 5 and Block 1 difference scores and sentence lengths were also statistically insignificant. For all subjects, the mean sentence length was 4.2 ($SD = 1.0$).

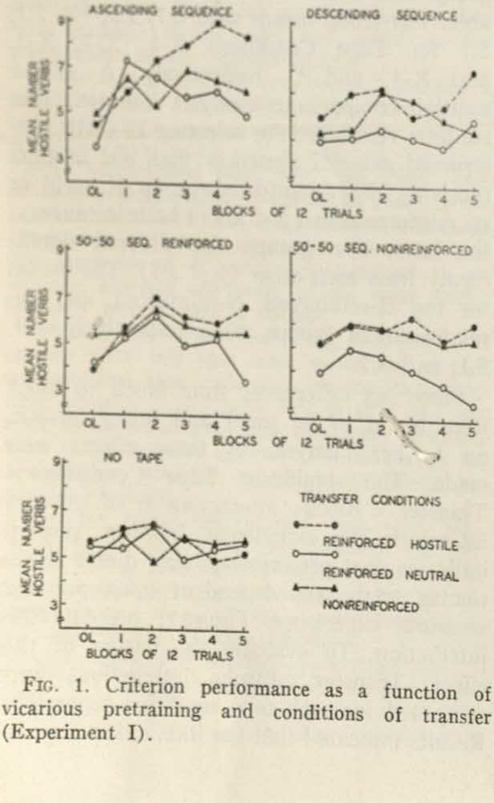


FIG. 1. Criterion performance as a function of vicarious pretraining and conditions of transfer (Experiment I).

EXPERIMENT II

The results of Experiment I showed that listening to a reinforced acquisition sequence of hostile responses increased the frequency of hostile verb choices early in transfer and facilitated subsequent conditioning under direct reinforcement. However, since the last block of the ascending sequence consisted solely of hostile verb choices, it is possible that hearing another child construct only hostile sentences immediately prior to transfer was of primary importance in producing the obtained transfer effects, rather than the acquisition sequence per se. Experiment II was designed to investigate this possibility by employing vicarious blocks consisting entirely of hostile sentences in either the initial or last positions of otherwise randomly constituted sequences.

An additional purpose of Experiment II was to determine whether conditionability would vary as a function of the pronoun class a subject was required to use in constructing sentences. In studying the pronoun-class variable, it was of interest (a) to examine whether the conditioning of hostility would be more difficult when subjects were required to make reference to themselves ("I") rather than to others ("they") in their sentence constructions, and (b) to determine whether vicarious exposure to expressions of hostility in the third person would facilitate the conditioning of hostility in the first person.

Method

Subjects. The subjects were 72 male and 87 female seventh, eighth, and ninth graders. While the assignment of subjects to the experimental conditions was random, males and females were approximately equally represented in each of the subgroups.

Stimulus materials. The stimulus materials were the same as those used in Experiment I. However, in addition to utilizing the 60 index cards containing H-N verb pairs and the pronoun "they," a second set of 60 index cards was constructed containing the same verb pairs and the pronoun "I." Tapes were recorded in the same way as in Experiment I, with a female graduate student serving as the experimenter and a male junior high school student as the subject.

Procedure. During the initial instructions and the 12 operant-level trials all subjects were treated identically to subjects in Experiment I. Following the operant-level trials, four experimental groups were given the same instructions as those given to subjects in Experiment I. For subjects in Group XR-5, the first 48 vicarious trials consisted of a random sequence of 18 H and 30 N sentences with the restriction that there be either 4 or 5 H sentences in each of the 4 blocks of 12 trials. The fifth block consisted of 12 H sentences. All of the H sentences were verbally reinforced on tape by the experimenter saying "good." The subjects in Group XN-5 listened to this same tape sequence but received no vicarious reinforcement. The subjects in Groups XR-1 and XN-1 received the same treatment as subjects in Groups XR-5 and XN-5, respectively, except that the block consisting entirely of H sentences was presented as the first block on the tapes. A fifth group of subjects (Group C) did not listen to any of the tapes and was asked to add columns of numbers for the length of time it took to play the tapes for the experimental groups.

The procedures during the transfer phase, as well as the conditions of transfer, were identical to those of Experiment I with the exception that for half of the subjects in each group the index cards contained the pronoun "I." There were 30 subjects in each of the four X groups, with 5 subjects in each of the six transfer-condition-pronoun subgroups. Group C contained 39 subjects, 7 subjects in each of the transfer conditions utilizing the pronoun "they," and 6 subjects in each of the transfer conditions utilizing the pronoun "I."

The same questionnaire as the one employed in Experiment I was administered to the subjects at the completion of the transfer phase. However, the following additional question was inserted after Question 1: "While going through the cards, did you think that you were supposed to make up your sentences in any particular way? In what way?"

RESULTS

As in Experiment I, the criterion measure was the number of hostile verb choices per block of 12 trials.

Group C Results

Figure 2 presents the mean performance over blocks for the subgroups receiving no vicarious experience. An analysis of variance on the operant-level data failed to reveal any statistically significant effects, thus indicating the comparability of the subgroups prior to the introduction of the experimental treatments. A 3 (Transfer) × 2 (Pronouns) analysis of covariance on Block 1 scores, adjusted for operant-level differences, likewise failed to reveal any reliable subgroup differences.

A Transfer × Pronouns × Blocks analysis of variance indicated that those subjects who were required to start their sentences with "I" used significantly fewer H verbs than subjects required to start their sentences with "they," $F\ (1,\ 33) = 4.47$, $p < .05$. The mean number of H-verb choices for the pronouns "I" and "they" was 3.9 and 4.7, respectively. Since the triple interaction was also statistically significant, $F\ (8,\ 132) = 2.38$, $p < .05$, the Transfer × Blocks interaction was computed for each pronoun condition separately. For the pronoun "I," this interaction was not reliabile ($F < 1$). When "they" was used as the pronoun, however, a statistically significant Transfer × Blocks interaction was obtained, $F\ (8,\ 72) = 3.00$, $p < .01$. Trend analyses revealed that subjects reinforced for choice of H verbs increased their use of H verbs over blocks, and subjects receiving no reinforcement decreased their use of H verbs ($p < .01$ for linear trend in both cases). Reinforcement for N verbs did not result in a statistically reliable trend.

Experimental Group Results

A 2 (Position of H Block) × 2 (Presence or Absence of Vicarious Reinforcement) × 3 (Transfer) × 2 (Pronouns) analysis of variance on the operant level data revealed that the Reinforcement × H-Block Position × Pronouns interaction, $F\ (1,\ 96) = 5.07$,

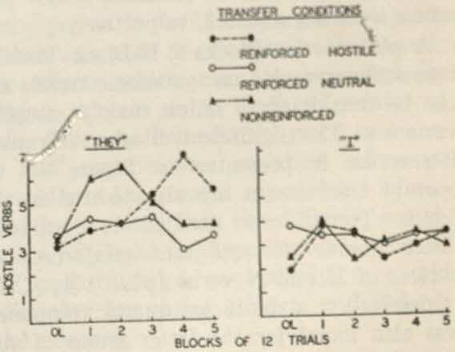


Fig. 2. Criterion performance of control subjects under the pronouns "they" and "I" (Experiment II).

$p < .05$, was the only statistically significant source of variance. This triple interaction is due to a Reinforcement × H-Block Position interaction in the data of those subjects subsequently conditioned under the pronoun "they." The operant level means for Groups XR-5, XR-1, XN-5, and XN-1 were 4.3, 2.7, 3.6, and 5.0, respectively. For subjects subsequently conditioned under the pronoun "I," the X-group means were similar and averaged 3.4.

An analysis of covariance on Block 1 scores revealed that the main effect for pronouns, F (1, 95) = 10.28, $p < .01$, was the only significant source of variance. The adjusted means for the pronouns "they" and "I" were 5.1 and 3.9, respectively.

Since reliable operant level differences were obtained, a 2 × 2 × 3 × 2 × 5 (blocks) analysis of covariance was conducted on transfer performance. This analysis revealed significant main effects for transfer, F (2, 95) = 12.02, $p < .001$, and pronouns, F (1, 95) = 33.22, $p < .001$; the following interactions were also found to be significant: Blocks × H-Block Position, F (4, 384) = 2.83, $p < .05$, and Blocks × Transfer, F (8, 384) = 2.76, $p < .01$. The results of t tests on the adjusted means indicated that subjects reinforced for selecting hostile verbs ($M = 5.3$) chose significantly more hostile verbs than subjects reinforced for choosing either neutral verbs ($M = 3.9$, $p < .001$) or subjects receiving no reinforcement ($M = 4.5$, $p < .01$); the latter two groups also differed significantly from each other ($p < .05$). The adjusted means for the "I" and "they" pronouns were 3.9 and 5.2, respectively.

A plot of the Blocks × H-Block Position interaction revealed no systematic trends and can be considered to reflect mainly sampling variation. The significant Blocks × Transfer interaction is presented in Figure 3; the operant level means are also plotted in this figure. Trend tests resulted in significant linear trends for subjects reinforced for choices of H and N verbs (p's $< .01$, $< .05$, respectively); a significant quartic component was also found for the latter group of subjects ($p < .01$). No reliable trend was found for subjects receiving no reinforcement for their sentence constructions.

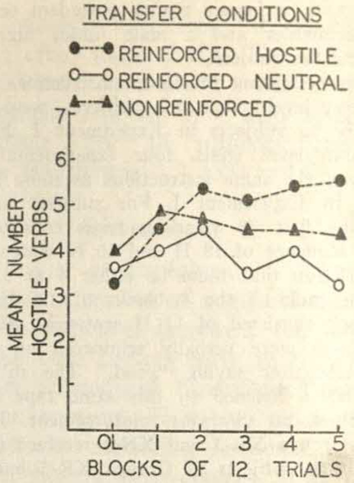


FIG. 3. Criterion performance of experimental subjects collapsed over conditions of vicarious pretraining and pronoun class (Experiment II).

To determine whether the obtained increases in hostile verb choices under the pronoun "they" differed for those experimental and control subgroups reinforced for hostile verb choices, a Groups × Blocks analysis of variance was conducted. Results indicated that both the main effect for groups, F (1, 25) = 1.75, $p > .05$, and the Groups × Blocks interaction, F (4, 100) = 1.95, $p > .05$, were not statistically significant. A similar analysis for the experimental and control subgroups reinforced for neutral verbs in transfer also yielded nonsignificant ratios for groups, F (1, 25) = 2.27, $p > .05$, and for the Groups × Blocks interaction, F (4, 100) = 1.05, $p > .05$.

Awareness Results

Two students, neither of whom served as the experimenter in this study, independently rated the postexperimental questionnaire protocols for "awareness." Criteria were the same as those employed in Experiment I. Twenty-six of the 159 subjects were classified as "aware" by both raters, there being only two disagreements.

The distribution of the number of aware subjects over the vicarious treatment groups is presented in Table 1. Performance curves were plotted separately for those aware and unaware subjects who were reinforced for choosing hostile and neutral verbs in trans-

TABLE 1

NUMBER OF SUBJECTS CLASSIFIED AS AWARE IN THE SUBGROUPS OF EXPERIMENT II

Groups	Transfer conditions					
	Reinforced H		Reinforced N		Nonreinforced	
	They	I	They	I	They	I
XR-5		2	1	2		1
XN-5	3	1	2	2		1
XR-1		1				
XN-1	3			2		
C	2		2	1		

fer. No systematic performance differences attributable to awareness were evident.

DISCUSSION

The present results indicate that the vicarious scheduling of sentences employing hostile and neutral verbs critically influences the subsequent conditioning of these same verb classes. In Experiment I it was found that an experimenter who failed to obtain conditioning when no vicarious pretraining was given could effectively condition hostile and neutral verb choices in subjects who previously had heard the experimenter reinforce another child for progressive increases in the use of hostile verbs (ascending sequence group). Yet, independent of the presence or absence of vicarious reinforcement, when the model failed to modify his behavior and emitted sentences in essentially "random" fashion (50-50 sequence groups), only sentences containing neutral verbs could be conditioned. That conditionability for hostile verbs over the first two transfer blocks was essentially as great for subjects in the 50-50 reinforcement groups as for subjects in the ascending sequence groups (see Figure 1), however, suggests that vicariously reinforced random schedules may also facilitate the direct conditioning of hostility.

In Experiment II, subjects who received vicarious pretraining showed increases in the use of both hostile and neutral verbs under direct-reinforcement conditions. However, these increases differed reliably from control-group performance only when "I" was used as the pronoun in transfer. Inasmuch as the control subgroups failed to show significant increases in either hostile or neutral verb choices under the first-person pronoun, these results, in agreement with the trends noted in Experiment I, indicate that prior exposure to vicarious random schedules will facilitate the conditioning of both verb classes. Moreover, consistent with the results of other observational learning studies employing both grade-school children (Bandura & Walters, 1963; Simon, Ditrichs, & Jamison, 1965) and adults (Marston, 1966), this facilitation is obtained whether or not the model receives reinforcement for his responses. However, it should be noted that the failure of vicarious reinforcement to *differentially* facilitate vicarious learning in Experiment II is contrary to results obtained with adult subjects (Kanfer & Marston, 1963; Marston, 1966) and suggests that vicarious reinforcement may play a lesser role in modeling behavior for subjects of younger ages.

The failure of vicarious pretraining to facilitate the conditioning of hostile and neutral verb choices under the pronoun "they" in Experiment II seems attributable both to the nature of the vicarious schedules utilized in this experiment as well as to the relative effectiveness of the experimenter employed as a reinforcing agent. Concerning the nature of the schedules employed in Experiments I and II, it is clear that with respect to both the immediate and prolonged transfer effects of vicarious pretraining, subjects exposed to the vicarious acquisition sequence in Experiment I showed the most pronounced conditionability for hostile verb choices. This suggests the importance of the acquisition sequence per se in facilitating conditionability and indicates that this facilitation transcends those effects which may be produced by exposing subjects to a block of reinforced hostile sentences immediately prior to transfer. That experimenter variables influenced the results of the present experiments is suggested by the fact that the different experimenters employed in Experiments I and II failed and succeeded, respectively, to obtain conditioning in control subjects with "they" as the pronoun.

To account for the enhanced conditionability that was found in Experiments I and II following vicarious pretraining, it seems plausible to suggest that listening to a model construct hostile sentences aided the subjects in identifying hostility as relevant to the task. Subsequent verbal reinforcement for nonhostile sentence constructions then leads to significant increases in the frequency of neutral verb choices. However, presumably due to prior learned inhibitions about expression of hostility, subsequent verbal reinforcement for hostile sentence constructions generally results in no or only limited increases in the choice of hostile verbs. Nonetheless, the relatively marked conditioning effect displayed by subjects in the ascending sequence group suggests that these learned inhibitions for hostile responding can be overcome by appropriate vicarious pretraining schedules, and that an important condition for achieving this effect is for the model to *noticeably* modify his behavior towards increasing expressions of hostility during the vicarious session. Moreover, the present results indicate that, within the limits of the number of transfer trials employed, the absence of direct reinforcement following exposure to the vicarious ascending schedule results in no further increases in the frequency of hostile verb choices.

The failure to obtain conditioning following the vicarious descending schedule appears consistent with results recently reported by Marston (1964). While a number of procedural differences exist between the latter study and the present experiments, it is of interest to note that Marston found that a vicarious descending sequence, reinforced or not, produces extinction following acquisition with vicarious reinforcement. The present results supplement this finding by indicating that direct acquisition is not facilitated by prior exposure to a vicarious descending sequence.

The influence of awareness could not be assessed in Experiment I since only two subjects verbalized the response-reinforcement contingency. According to the criteria for awareness utilized in Experiment II, 23% of the subjects receiving direct reinforcement in transfer were classified as aware. However, since awareness was not related to conditionability, the finding that different defining operations yielded different percentages of aware subjects is of little importance in itself.

It has been suggested (e.g., Portnoy & Salzinger, 1964) that conditionability may vary inversely with sentence length because of a differential delay of reinforcement for the correct response. No such relationship was observed in Experiment I. Certainly, a more adequate test of the suggested dependency would be provided by a design in which sentence length is treated as a main effect (perhaps by the use of appropriate instructions to the subjects).

REFERENCES

Bandura, A., & Walters, R. H. *Social learning and personality development.* New York: Holt, 1963.

Fadigan, J. Indirect verbal conditioning. Unpublished master's thesis, Florida State University, 1961.

Greenspoon, J. The effect of a verbal stimulus as a reinforcement. *Proceedings of the Indiana Academy of Sciences,* 1950, **59**, 287.

Hurlock, R. E., & McCain, G. Vicarious acquisition and vicarious extinction in an operant verbal conditioning situation. Paper read at Midwestern Psychological Association, St. Louis, May 1964.

Kanfer, F. H., & Marston, A. R. Human reinforcement: Vicarious and direct. *Journal of Experimental Psychology,* 1963, **65**, 292–296.

Marlowe, D., Beecher, R. S., Cook, J. B., & Doob, A. N. The approval motive, vicarious reinforcement and verbal conditioning. *Perceptual and Motor Skills,* 1964, **19**, 523–530.

Marston, A. R. Variables in extinction following acquisition with vicarious reinforcement. *Journal of Experimental Psychology,* 1964, **68**, 312–315.

Marston, A. R. Determinants of the effects of vicarious reinforcement. *Journal of Experimental Psychology,* 1966, **71**, 550–558.

Marston, A. R., & Kanfer, F. H. Group size and number of vicarious reinforcements in verbal learning. *Journal of Experimental Psychology,* 1963, **65**, 593–596.

Portnoy, S., & Salzinger, K. The conditionability of different verbal response classes: Positive, negative and nonaffect statements. *Journal of General Psychology,* 1964, **70**, 311–323.

Simon, S., Ditrichs, R., & Jamison, N. Vicarious learning of common and uncommon associations in children. *Psychonomic Science,* 1965, **3**, 345–346.

(Received August 20, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 79–85

BRIEF ARTICLES

SITUATIONAL DETERMINANTS OF AFFILIATIVE PREFERENCE UNDER STRESS[1]

ROBERT L. HELMREICH AND BARRY E. COLLINS
University of Texas *University of California, Los Angeles*

The study tests the hypothesis that affiliation under stress might be produced by a dependency-motivation mechanism—rather than the social comparison and direct-anxiety-reduction mechanisms discussed in previous papers. Each S was given the opportunity to wait (or work) in a leader-dominated group as well as the standard opportunities to wait alone or in a group of peers. Ss facing a fearful situation showed a significant decrease in preference for the company of peers in favor of being in a leader-dominated group, supporting the dependency-motivation hypothesis. Ss initially run and stressed in groups evidenced markedly greater affiliative choices (leader plus peer) than Ss run alone. When choices were made for working under stress, no birth-order effects were found; when choices were made for waiting for stress, the previously reported birth-order interaction was replicated.

Frightened subjects have consistently shown an increased desire for the company of others while awaiting a threatening event. Schachter (1959) developed what has become a standard experimental paradigm by announcing to subjects that they would receive painful electric shock and then offering them the opportunity to wait for the shock alone or with others. Other investigators (Gerard & Rabbie, 1961; Miller & Zimbardo, in press; Rabbie, 1963; Zimbardo & Formica, 1963) have used this paradigm and found similar increases in affiliative desires under high fear. A number of studies (Darley & Aronson, 1966; Gerard & Rabbie, 1961; Schachter, 1959; Zimbardo & Formica, 1963) have reported that stress-induced affiliative needs are greater among firstborn and only children than among later borns.

Schachter (1959) suggested that two mechanisms may be responsible for affiliative responses:

1. Individuals may need others for evaluation of their own emotional reactions. Persons in the same emotional state can provide needed information on what responses are appropriate to ambiguous arousal cues. (See Festinger's 1954 theory of social comparison processes and Schachter's 1964 application to cognitive determination of emotion.)

2. The company of others may also provide a direct reduction of the anxiety which the subject feels and hence provide a motivation for affiliation. Wrightsman (1960) found the presence of others effective in reducing anxiety, while Darley and Aronson (1966) found both motives operating to stimulate affiliation.

Janis (1963) discussed another mechanism which may explain the increased importance of primary groups and especially authority figures such as leaders in times of stress. He pointed to stress-induced dependency reactions towards authority figures and suggested that the increased dependency toward authority occurs because the authority can serve as a "danger control authority" which can reduce the severity of threat perceived by the individual. Thus, the ability to establish a dependent relationship could provide another strong motivation for affiliation under fear.

A presently unanswered question of considerable theoretical and practical importance is whether the increased desire for affiliation which has been previously observed for persons who were *awaiting* a frightening event would also be observed among persons actually facing a threat. For example, would the conclusions imply a need for affiliation during combat or just while waiting to go into battle? It seems plausible that there may be a qualitative difference between gregariousness while anticipating threat and while working in a stressful situation.

Overview of the Research

The research reported here consists of two studies, designed to clarify some of the questions raised in earlier investigations of affiliative preference under stress. In Study I, subjects under high or low fear were offered the chance to remain alone, to be with a group of peers in the same emotional state, or to be in a group with a leader. The affiliative preferences were measured for a *working* situation—subjects were asked to choose with whom they would like to work while receiving stimulation. Contrary to earlier research, the results showed a clear-cut decrease in preference for the company of peers under high fear. Also, unlike earlier investigations of fear-produced affiliation, no differential effects for birth order were found. Study II was undertaken to clarify and refine these results. In this experiment subjects were offered the same choices for affiliation employed in Study I, and these choices were offered both for the *working* condition and for the more typical situation of *waiting* for shock. The only other difference in the two studies was that subjects in Study II were run in groups rather than individually as in Study I. The decrease in preference for the company of peers in the working condition found in Study I was replicated, while a different pattern of results was obtained for the waiting condition. An overall difference in desire to affiliate was also found in Study II, suggesting the possibility that whether subjects are run in groups or individually is an important variable.

---

[1] Portions of this study were based on a paper read at Eastern Psychological Association meetings, New York, April 1966. The research was supported by National Science Foundation Grant GS 492 to the junior author.

STUDY I

Method

Subjects. Sixty male high school students between the ages of 14 and 18 were recruited through an advertisement in the local newspaper for a psychological study. From the list of those replying to the advertisement, 30 firstborns and 30 later borns were selected randomly and used in the study. Within each group, subjects were assigned randomly to one of the three conditions of fear arousal employed.

Procedure. After arriving for the experiment, the subjects were conducted alone to an office furnished with a desk, chairs, and dummy electronic equipment. Each subject was told that the "doctor" would be in shortly. The experimenter then entered the room and introduced himself as a doctor studying the physiology of mental processes. The experiment was described as an investigation of the physiological reactions occurring during mental activity. The experimenter stated that a basal rate for each subject would be determined by using the electrode present to take readings on a "master polygraph recorder."

A laboratory assistant then arrived and attached a standard skin electrode to the subject's arm. The assistant gave the subject instructions for filling out a Family Background Inventory, Form A of the Sixteen Personality Factor Questionnaire (Cattell's 16 PF—Cattell & Eber, 1964) and a 33-question multiple-choice measure of self-esteem (feelings of social inadequacy and social inhibitions) taken from Janis and Field (1959).

When the subject completed the forms (this took between 45 minutes and an hour), the assistant entered the room, unhooked the electrode, and told the subject that the doctor would be in to explain the second part of the study.

The experimenter then informed the subject that the second part would take place in a different laboratory and would consist of an investigation of problem-solving behavior. The experimenter remarked that basal rates had been obtained and that two electrodes would be attached to the subject during Part II—one a recording electrode, the other an electrode for stimulation. At this point, the fear manipulations were introduced.

Low-fear subjects were told that they would receive subliminal electric stimulations during Part II. It was emphasized that they would not feel the stimulation, and the word "shock" was never used. The experimenter said that he was interested in the effects of stimulation on performance. *Moderate-fear* subjects were told that they would receive several electric shocks which were described as being noticeable but not unpleasant. The experimenter remarked that he was interested in the effects of shock on peformance. *High-fear* subjects were told that they would receive strong and painful electric shocks and that blood samples would be drawn to measure any changes in blood composition caused by the experimental procedure. The experimenter also stated that he was interested in the effects of shock on performance. All subjects were then told that part of the study was concerned with investigating differences between the reactions of people working alone and those working in groups. He added that a number of subjects were just finishing Part I, and the subject could have a choice of the condition to which he would be assigned for Part II.

Next, the three choices were described. The *alone* condition was described as a situation in which the subjects would work on problems by himself in a lab room. The *informal group* condition (peer group) was depicted as consisting of subjects working individually on problems in the company of three other subjects receiving the same stimulation. The point was clearly made that the subject would be working by himself but would be in the company of same-state peers. The *leader group* was described as consisting of two other high school subjects and a graduate-student leader. The experimenter said that this group would work as a unit on the problems, and that the leader would assign tasks, direct performance, and control the group's activity. The subject was then asked to indicate his preference on a sheet of paper. Since subjects were run individually, no possibility existed for the choice to be affected by acquaintance with other subjects. No subject asked to drop out of the study before Part II. The experimenter then explained the real nature of the experiment and answered questions. All the subjects were then enjoined to secrecy.

Results and Discussion

The effectiveness of the fear manipulation was evaluated by means of an item on the final questionnaire asking the subjects to report how nervous and ill at ease they felt concerning the stimulation they were about to receive (on a

TABLE 1

STUDY I: MEAN RATINGS OF FEAR[a]

	Low fear	Moderate fear	High fear
Firstborn	9.87	19.10	38.50
Later born	$10.43_{a, c, d}$	18.84_{b}	$36.75_{b, c, d}$

Note.—Numbers with no subscripts in common are nonsignificant ($p > 10$). Numbers with one subscript in common are significant at $p < .05$. Numbers with two subscripts in common are significant at $p < .01$.

[a] Significance levels are for contrasts between combined means of firstborns and later borns, $df = 1/42$ for all comparisons. No differences between firstborns and later borns approached significance.

scale running from 0 to 100). The differences among the three fear conditions were in the expected direction, and the main effect ($F = 8.18$, $df = 2/42$) was significant at better than the .01 level. Comparisons among the individual condition means revealed that high-fear subjects were significantly more nervous than moderate-fear ($p < .05$) and low-fear ($p < .01$) subjects, while moderate-fear subjects did not differ significantly from low-fear subjects at even the .10 level. These results are summarized in Table 1. No significant birth-order effects were found for reported fear.

The choices of affiliation revealed a significant *decrease* in preference for the company of peers in the high-fear group contrasted with the low- and moderate-fear conditions (Table 2).

TABLE 2

STUDY I: DIRECTION OF AFFILIATION × FEAR CONDITION × BIRTH ORDER

	% alone choice	% peer-group choice	% leader-group choice	N
	Low-fear condition			
Firstborn	30	40	30	10
Later born	50	40	10	10
Firstborn and later born combined	40	40	20	20
	Moderate-fear condition			
Firstborn	40	40	20	10
Later born	40	50	10	10
Firstborn and later born combined	40	45	15	20
	High-fear condition			
Firstborn	50	10	40	10
Later born	60	10	30	10
Firstborn and later born combined	55	10	35	20

TABLE 3

STUDY I: DIRECTION OF AFFILIATION[a]

	% alone choice	% peer-group choice	% leader-group choice	N
	Low- and moderate-fear condition			
Firstborn	35	40	25	20
Later born	45	45	10	20
Firstborn and later born combined	40	42.5	17.5	40
	High-fear condition			
Firstborn	50	10	40	10
Later born	60	10	30	10
Firstborn and later born combined	55	10	35	20

[a] Low- and moderate-fear groups combined.

Since no differences in affiliative preference and no significant differences in level of fear were found between low- and moderate-fear conditions, these two groups were combined for analysis. A 2 × 3 chi-square analysis of affiliative choices (Table 3) indicates that more subjects chose both the alone and leader groups under high fear ($\chi^2 = 9.97$, $df = 2$, $p < .01$). The a priori hypothesis had predicted only the increase in preference for the *leader group* under high fear and not the increased preference for the *alone* choice which contributes to the significance of the chi-square in Table 3. However, the increase in preference for the leader group is significant independently, as reflected by a 2 × 2 chi-square which combines the *alone* and *peer* columns of Table 3 ($\chi^2 = 9.98$, $p < .005$, $df = 1$).

Contrary to most earlier studies of fear-induced affiliation, no increase in preference for companionship over solitude was found under high fear. The chi-square combining the peer and alone columns of Table 3 clearly was not significant ($\chi^2 = 1.20$, $df = 1$).

No differences in affiliative preferences between firstborns and later borns were found. Both made each affiliative choice with almost identical frequencies (see Table 2).

Results from the personality measures were disappointing. Although 6 of 56 computed Fs were significant, the obtained results were generally uninteresting.[2] In particular, none of the personality indexes reputed to measure self-esteem or dependency was significantly related to affiliative choice.

---

[2] A complete report of the personality findings from Study I may be found in a report, "Decrease in Preference for the Company of Peers under High Fear," by Robert L. Helmreich, deposited with the American Documentation Institute, Washington, D. C.

The decrease in preference for the company of peers as well as the failure to find an absolute increase in preference for companionship under high fear are counter to behavior reported in most fear-affiliation studies. On the basis of this study, it is impossible to determine whether the obtained findings diverge from previous results because (*a*) this is the first study to offer the leader choice, with its opportunity for dependency, or (*b*) this study offered affiliative choices for a working situation rather than the waiting for pain situations used in previous studies.

The possibility also existed that the high preference for solitude in this study resulted from a strong "set" for solitude engendered in subjects who had been working alone for a long period prior to the affiliative choice. This set effect might also have been intensified by the fear manipulation, causing the unexpected increase in preference for working alone under high fear.

The second study was designed to clarify these issues by replicating Study I and, in addition, by offering the same three affiliative choices for a waiting period prior to shock as well as for working in a threatening situation.

STUDY II

Method

Subjects. Subjects were 120 male high school students between the ages of 15 and 18 recruited through a local high school.

Procedure. All subjects arrived simultaneously for the study and were assigned randomly to experimental conditions by receptionists. Each subject was sent to one of two large classrooms. All subjects in one room were assigned to the high-fear condition, while all of those in the other room were assigned to the low-fear condition.

After all subjects had arrived, an experimenter introduced himself and stated that a number of studies were being conducted simultaneously during the morning. He added that the several portions of the experiment would be explained by the persons in charge of the particular studies. The subjects were initially asked to fill out a background questionnaire giving biographical data and ordinal position and a group intelligence test (the Wonderlic Personnel Test, Form A). After all subjects had completed these forms the fear manipulations were introduced.

Subjects in each fear condition were randomly assigned to 1 of 12 groups. Six graduate-student experimenters took groups of five subjects to a smaller classroom at which time the fear manipulation was administered to these groups of five. Each experimenter ran either four high-fear groups or four low-fear groups. Two of the groups run by each of the experimenters were in the working condition and two in the waiting condition. The order of groups for each experimenter was randomly determined.

Fear manipulation. All low-fear subjects were told that the experimenters were interested in measuring physiological changes while persons engaged in mental activity. They were told that they would solve simple problems while receiving subliminal stimulation and while recordings were made of changes in the activity of the nervous system, and that a polygraph recorder would be hooked to an electrode attached to their arm. The experimenter explained that recordings would be made while they solved problems.

High-fear subjects were told that the experimenters were interested in measuring physiological changes while a person was engaged in mental activity and received electrical stimulation. The subjects were told that a recording electrode would be attached to them and that a stimulating electrode would also be connected, and that they would receive strong shocks which most subjects found painful. In addition, they were informed that changes in content of blood during mental activity and stimulation would be measured. The experimenter added that a nurse would use a hypodermic needle to take three blood samples—at the beginning, in the middle, and at the end of the experimental task. The experimenter concluded, "Remember, you will receive several strong electric shocks and three blood samples will be taken." At this point the experimenters showed the subjects a large hypodermic needle and stated that a similar needle would be used to withdraw blood.

Affiliative choices. Working-condition subjects were informed that the social setting present while persons solved tasks and received stimulation could have strong effects on the nature of the results. Therefore, subjects were permitted to decide for themselves what social setting they would work in. The subjects were offered three choices for working and receiving stimulation—to work alone in a private room, to work in the company of three other high school students receiving similar stimulation, or to work in a "leader group" with two other high school subjects and a college-graduate student serving as group leader. The experimenter added that in the leader group the students worked together on problems, and the graduate student supervised and controlled group activity.

Waiting-condition subjects were told that the presence of other people could affect physiological reactions and that what a person does just before taking part in a study can have a big effect on his reactions. Therefore, the experimenter explained, subjects could decide for themselves how they would wait for the task to begin while the equipment was readied. Three choices were described—to wait alone in a private room, to wait in an informal group with three other subjects, or to wait in a leader group with two other high school students and a college-graduate student who would supervise what the subjects did during the waiting period. The experimenter then passed out a questionnaire asking subjects to indicate their preference for each of the three possible choices. The subjects indicated their preference between each of the three paired comparisons possible among the three conditions. Each preference was rated on a 4-point scale running from "very much prefer this choice" through "slightly prefer" to "very much prefer" for the other choice on each comparison.

After all subjects in each group completed the affiliative choice questionnaire, it was returned to the classroom by the experimenter.

A 33-item Nowlis Adjective Check List of Mood was administered as a "take measure" on the fear manipulation. Because of time limitations in the study, the design called for only 80 of the 120 subjects to complete the mood-adjective check list.

After taking part in a second and separate study, all subjects were debriefed, and the purposes of the experiment were fully explained.

Results and Discussion

Set effects. Perhaps the most striking difference between Studies I and II is that, while almost half of the subjects in Study I chose solitude over companionship, more than 90% of the subjects in Study II could be classified as having chosen to be with either the peer group or the leader group. This was true in both the waiting and the working conditions. Thus, desire for affiliation was clearly stronger among subjects who had spent a considerable amount of time in groups prior to the affiliative choice than it was among those who had been alone before being offered a chance to affiliate. It appears that the prior social setting can have a strong effect on the desire for gregariousness.

The subject population was extremely heterogeneous. Results of the Wonderlic Personnel Test converted to high school norms indicated that the population ranged from the first to the ninety-ninth percentile in intelligence. On the basis of an arbitrary a priori decision, it was assumed that subjects scoring in the first percentile or below would be unable to understand the experimental manipulations. Therefore, all subjects with absolute scores below 8 (first percentile) on the Wonderlic were eliminated from further analysis. This resulted in a loss of 12 subjects. Of the remaining 108 subjects, 54 were firstborn and 54 later born (a chance result, as subjects had not been recruited on the basis of ordinal position).

Effectiveness of fear. The first 80 subjects who received the fear manipulation filled out the 33-item mood-adjective check list immediately after receiving the fear manipulation. With 5 of these 80 subjects dropped because of low IQ, 40 low-fear and 35 high-fear subjects were used in the analysis for effectiveness of fear. On an a priori basis, 6 of the 33 adjective pairs were selected to form a "fear index" (anxious-fearful; disturbed-upset; easygoing-relaxed, negatively scored; jittery-nervous; frightened-worried; quiet-peaceful, negatively scored). A 2 × 2 (Fear × Ordinal Position) unweighted-means analysis of variance was performed on the summed fear-index score. The main effect for fear was significant at better than the .001 level ($F = 14.70$, $df = 1/71$), demonstrating that the manipulation was highly effective in making the high-fear group more nervous and apprehensive than the low-fear subjects.[3] A trend toward an interaction between fear and ordinal position ($p < .10$) was found, with firstborns less frightened than later borns in the low-fear condition, but more apprehensive under high fear. Fear means were low fear—firstborns, 5.41; later borns, 6.30; high fear—firstborns, 11.64; later borns, 8.62. The analyses of variance are summarized in Table 4. A similar trend was observed in Study I.

TABLE 4

STUDY II: ANALYSIS OF VARIANCE FOR FEAR INDEX[a]

	F
Fear (A)	14.70***
Ordinal position (B)	0.91
A × B	3.08*

[a] Posttest only; $df = 1, 71$ for all analyses.
* $p < .10$.
*** $p < .001$.

Affiliative preference. A three-way unweighted-means analysis of variance (Winer, 1965, p. 602) was performed (Fear × Working-Waiting × Firstborn-Later Born) on the ratings which contrasted each of the three possible comparisons among the three affiliative choices. A significant main effect for fear was found for the leader-versus peer-group ratios ($F = 4.98$, $p = .05$, $df = 1/100$). This replicates Study I, as the main effect reflects a *decrease* in preference for the peer group in favor of the leader group. A significant interaction ($F = 4.34$, $p < .05$, $df = 1/100$) between fear and ordinal position was also obtained for this comparison. As can be seen in Figure 1, all subjects in the working condition and later borns in the waiting condition showed *increased* desire for the leader group under high fear. Firstborns in the high-fear–waiting condition showed a slight decrease in preference for the leader group (as compared to the peer group). This finding *for the waiting condition* parallels earlier affiliation studies where firstborns have shown the greatest desire for the company of peers under high fear. Results of the analyses of variance for all choices are summarized in Table 5.

No significant main effects were found for the comparison between peer group and alone choices; however, a significant interaction be-

---

[3] Individual t tests indicated that each of the six adjective pairs discriminated between high- and low-fear conditions at better than the .05 level.

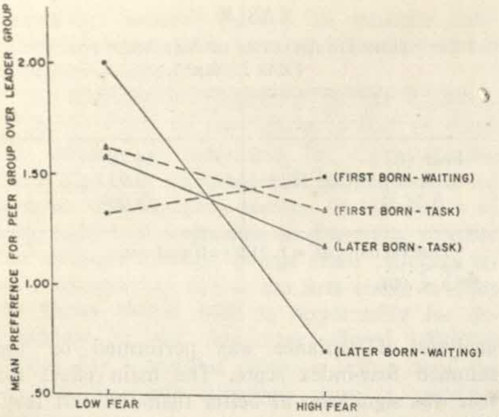


FIG. 1. Mean ratings on scale for peer-group versus leader-group comparison.

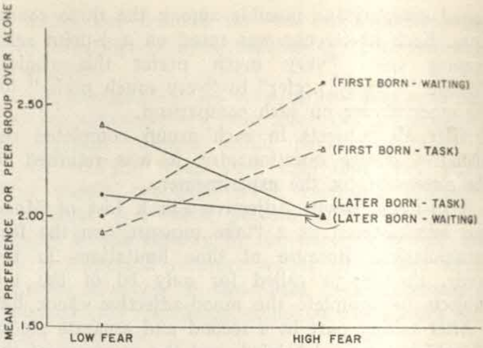


FIG. 2. Mean ratings on scale for peer-group versus alone comparison.

tween fear and ordinal position was found in this analysis. As can be seen in Figure 2, this interaction reflects an increased desire for affiliation with peers on the part of firstborns under high fear when the choice offered is solitude or peer companionship (i.e., the only choices offered in previous studies). This result, which held for both the waiting and task conditions, is similar to the changes in affiliative preference reported in earlier studies (e.g., Schachter, 1959), but it is opposite to the trend observed in Study I. It should be noted that subjects in Study I were run individually and allowed to choose only one of the three working conditions; subjects in Study II were run in groups and allowed to express a preference among each of the three possible pairings of the three affiliative conditions. No significant main effects or interactions were found for the contrast between the leader groups and alone choices.

Experimenter effects. Results were analyzed separately for subjects run by each experimenter. No significant effects were found for either level of fear or affiliative preference.

TABLE 5

STUDY II: ANALYSES OF VARIANCE FOR AFFILIATIVE CHOICE[a]

	F		
	Peer group vs. alone choice	Peer group vs. leader choice	Leader group vs. alone choice
Fear (A)	<1	4.98**	1.36
Working vs. waiting (B)	<1	<1	<1
Ordinal position (C)	<1	<1	<1
A × B	<1	<1	<1
A × C	4.87**	4.34**	<1
B × C	1.07	<1	<1
A × B × C	<1	2.62	1.18
MS (error)	0.71	0.99	0.64

[a] $df = 1{,}100$ for all analyses.
** $p < .05$.

IMPLICATIONS

Studies I and II indicate that offering the company of an older leader (who, according to the present theory, may satisfy dependency needs aroused by stress) can change the direction of affiliation under stress. Thus, it is established that dependency needs are at least one of the factors creating affiliative behavior under stress. It is difficult to assess from earlier studies what effect motivation for dependency may have had on direction of affiliation because dependency needs could contribute to the peer choice if the only other alternative were solitude.

The fact that the birth-order differences found in the waiting condition were not present for working choices may indicate a qualitative difference in the two conditions. Where the emphasis is on constructive performance, all subjects may perceive that the most satisfactory arrangement is to sacrifice autonomy and to accept a dependent relationship. In the waiting condition, on the other hand, firstborns may find social comparison with peers most important; later borns, who presumably have less need for social support, may find the greatest anxiety reduction through establishing a dependent relationship.

The striking increase in preference for the peer and leader groups under fear observed in Study II (as compared to Study I) is, of course, a between-experiment finding which could have been produced by any of the many differences between the two studies. Nonetheless, the hypothesis that the phenomenon can be attributed to the fact that subjects in Study I were run in solitude while subjects in Study II were run in groups adds another dimension to the assessment of affiliative preference. Subjects in Studies I and II had had significant experience either alone or in a group immediately prior to making an affiliative choice; that is, subjects spent between 40 minutes and an hour either alone or in a large group before being offered affiliative choices. In earlier studies (e.g., Schachter, 1959) subjects spent a short period of time in groups or alone and were then offered the choice of affiliation. It seems probable that in the present studies subjects acquired a set for solitude or gregariousness which is reflected in the choices made by low-fear subjects. The fear manipulation may have strengthened this set as the individual, faced with a threatening situation, might be reluctant to undertake any novel experience. Thus, individuals who had been alone might find a sudden confrontation with a group of people more threatening than solitude, while those who had been in groups might find being alone more threatening. There is a strong implication that one must control the subject's set, his situationally determined propensity for gregariousness, before one can compare the effects of different treatments or can speak with certainty about overall changes in motivation for affiliation.

REFERENCES

Cattell, R. S., & Eber, H. W. *Handbook for the Sixteen Personality Questionnaire.* Champaign, Ill.: Institute for Personality and Ability Testing, 1964.

Darley, J. M., & Aronson, E. Self-evaluation vs. direct anxiety reduction as determinants of the fear-affiliation relationship. *Journal of Experimental Social Psychology,* 1966, Suppl. 1, 66–79.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, 7, 117–140.

Gerard, H. B., & Rabbie, J. M. Fear and social comparisons. *Journal of Abnormal and Social Psychology,* 1961, 62, 586–592.

Janis, I. L. Group identification under conditions of external danger. *British Journal of Medical Psychology,* 1963, 36, 227–238.

Janis, I. L., & Field, P. B. Sex differences and personality factors related to persuasibility. In C. Hovland & I. L. Janis (Eds.), *Personality and persuasibility.* New Haven: Yale University Press, 1959. Pp. 55–68.

Miller, N., & Zimbardo, P. G. Motives for fear-induced affiliation: Emotional comparison or interpersonal similarity. *Journal of Personality,* in press.

Rabbie, J. M. Differential preference for companionship under threat. *Journal of Abnormal and Social Psychology,* 1963, 67, 643–648.

Schachter, S. *The psychology of affiliation.* Stanford: Stanford University Press, 1959.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Wrightsman, L. The effect of waiting with others on changes in level of felt anxiety. *Journal of Abnormal and Social Psychology,* 1960, 61, 216–222.

Zimbardo, P. G., & Formica, R. Emotional comparison and self-esteem as determinants of affiliation. *Journal of Personality,* 1963, 31, 141–162.

(Received April 20, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 85–91

THURSTONE'S MEASUREMENT OF SOCIAL VALUES REVISITED FORTY YEARS LATER[1]

CLYDE H. COOMBS[2]

University of Michigan

In 1927 Thurstone reported a scaling study of the judged seriousness of crimes and offenses ranging from rape to vagrancy. That study has been repeated 40 years later using the same stimuli, the same instructions, and an equivalent population of subjects. Findings include: (a) There was more homogeneity of attitudes toward the seriousness of crimes 40 years ago than today; (b) offenses against the person are now judged relatively more serious and sex offenses relatively less so; and (c) offenses against property, with the exception of arson, are judged less serious relative to the others than they were 40 years ago. Some sex differences are noted.

THURSTONE'S 1927 STUDY

In 1927, Thurstone published his law of comparative judgment (1927a), a theory and algorithm for measuring subjective magnitudes. The theory was developed in the context of psychophysical measurement, and Thurstone wished to demonstrate its possible usefulness in social psychology. To this end he made a number of applications to such areas as the measurement of attitudes, nationality preferences, and the judged seriousness of crimes and offenses.[3] Our interest

---

[1] This publication was supported in part by United States Public Health Service Grant MH 04236.

[2] I wish to thank the psychology sections and their instructors who were so cooperative and Frank M. Goode who supervised the analysis of the data.

[3] A posthumous collection of his papers (Thurstone, 1959).

TABLE 1

PROPORTION OF SUBJECTS WHO JUDGED THE ROW OFFENSE MORE SERIOUS THAN THE COLUMN OFFENSE

Offense	Abortion	Adultery	Arson	Assault & battery	Bootlegging	Burglary	Counterfeiting	Embezzlement	Forgery	Homicide	Kidnapping	Larceny	Libel	Perjury	Rape	Receiving stolen goods	Seduction	Smuggling	Vagrancy
Abortion		.46	.38	.44	.68	.53	.54	.57	.58	.11	.33	.50	.61	.58	.19	.66	.60	.78	.78
Adultery	.54		.36	.32	.71	.49	.53	.51	.57	.08	.21	.49	.56	.51	.10	.62	.60	.58	.82
Arson	.62	.64		.46	.95	.79	.86	.85	.89	.07	.29	.86	.85	.78	.19	.92	.70	.89	.96
Assault & battery	.56	.68	.54		.95	.81	.84	.85	.85	.06	.28	.85	.85	.78	.12	.93	.69	.88	.97
Bootlegging	.32	.29	.05	.05		.11	.12	.15	.18	.01	.02	.17	.28	.23	.02	.41	.25	.18	.78
Burglary	.47	.51	.21	.19	.89		.69	.57	.73	.03	.09	.64	.65	.57	.05	.90	.49	.76	.94
Counterfeiting	.46	.47	.14	.16	.88	.31		.39	.51	.02	.06	.43	.56	.44	.07	.85	.47	.68	.92
Embezzlement	.43	.49	.15	.15	.85	.43	.61		.67	.03	.08	.56	.66	.50	.08	.89	.51	.69	.94
Forgery	.42	.43	.11	.15	.82	.27	.49	.33		.02	.03	.42	.57	.40	.05	.78	.46	.58	.91
Homicide	.89	.92	.93	.94	.99	.97	.98	.97	.98		.89	.99	.99	.96	.73	.98	.94	.98	.98
Kidnapping	.67	.79	.71	.72	.98	.91	.94	.92	.97	.11		.94	.95	.90	.31	.98	.78	.96	.98
Larceny	.50	.51	.14	.15	.83	.36	.57	.44	.58	.01	.06		.68	.55	.07	.81	.54	.65	.94
Libel	.39	.44	.15	.15	.72	.35	.44	.34	.43	.01	.05	.32		.37	.06	.64	.44	.48	.86
Perjury	.42	.49	.22	.22	.77	.43	.56	.50	.60	.04	.10	.45	.63		.08	.72	.55	.57	.92
Rape	.81	.90	.81	.88	.98	.95	.93	.92	.95	.27	.69	.93	.94	.92		.96	.97	.94	.99
Receiving stolen goods	.34	.38	.08	.07	.59	.10	.15	.11	.22	.02	.02	.19	.36	.28	.04		.41	.19	.83
Seduction	.40	.40	.30	.31	.75	.51	.53	.49	.54	.06	.22	.46	.56	.45	.03	.59		.51	.79
Smuggling	.22	.42	.11	.12	.82	.24	.32	.31	.42	.02	.04	.35	.52	.43	.06	.81	.49		.93
Vagrancy	.22	.18	.04	.03	.22	.06	.08	.06	.09	.02	.02	.06	.14	.08	.01	.17	.21	.07	

here is in the last-mentioned study (Thurstone, 1927b) and the changes, 40 years later.

Thurstone utilized 266 University of Chicago students and presented them with 171 pairs of offenses (19 taken 2 at a time). His instructions to the subjects were as follows:

> The purpose of this study is to ascertain the opinions of several groups of people about crimes. The following list of crimes has been arranged in pairs. You will please decide which of each pair you think more serious and underline it.
>
> An example: Cheating—*Murder*.
>
> You would probably decide that murder is a more serious offense than cheating; therefore you would underline *Murder*.
>
> If you find a pair of crimes that seem equally serious, or equally inoffensive, be sure to underline one of them anyway, even if you have to make a sort of guess. Be sure to underline one in each pair [Thurstone, 1927b, p. 385].

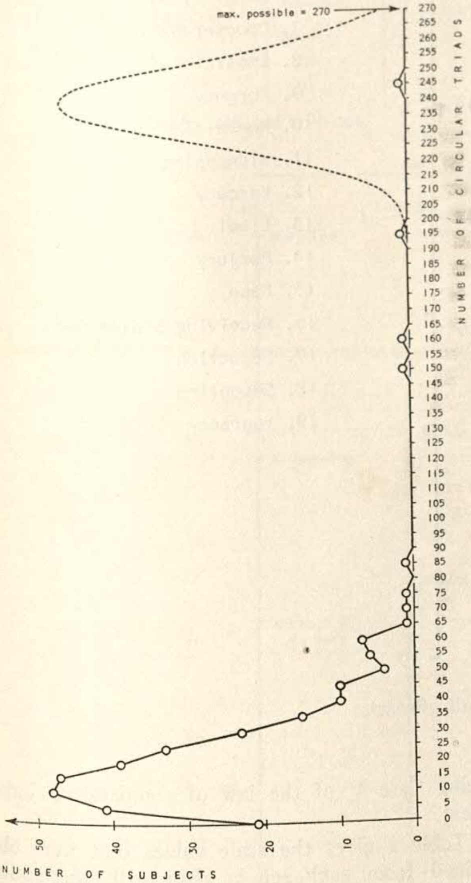


FIG. 1. Obtained frequency distribution of intransitivities (circles and solid lines), and a normal approximation to the expected distribution for random responses (dashed lines).

TABLE 2

CASE V SCALE VALUES OF JUDGED SERIOUSNESS OF OFFENSES

Offense	1966 data			1927 data
	Males	Females	Total	Thurstone
Abortion	40.0	50.8	45.62	69.3
Adultery	40.4	43.0	41.7	64.2
Arson	69.9	60.0	64.3	61.7
Assault & battery	61.7	65.5	63.6	45.0
Bootlegging	19.3	7.8	13.0	31.5
Burglary	45.0	45.6	45.2	46.1
Counterfeiting	41.8	33.6	37.4	49.9
Embezzlement	43.9	40.4	42.0	50.6
Forgery	37.1	30.6	33.5	47.7
Homicide	100.0	100.0	100.0	96.4
Kidnapping	77.9	79.9	79.1	67.1
Larceny	39.2	38.5	38.8	40.5
Libel	33.9	27.8	30.6	34.3
Perjury	42.1	38.1	39.8	51.2
Rape	87.2	86.1	86.6	100.0
Receiving stolen goods	18.9	16.4	17.4	30.5
Seduction	34.9	40.2	37.5	69.4
Smuggling	32.6	29.4	30.7	33.6
Vagrancy	0.0	0.0	0.0	0.0

1966 REPLICATION

The study was repeated, using the same stimuli and instructions, on 369 University of Michigan students in a number of sections of an elementary psychology course. An occasional student missed an item or a page in the list of pairs, but no pair had less than 363 respondents. The number of intransitive triples (circular triads) in each student's pair comparisons was an indication of the student's consistency, assuming that he was basically transitive except for random fluctuations. The frequency distribution[4] of circular triads is shown in Figure 1. If a student's choice on each pair were on a random basis with $p = 1/2$, the number of circular triads to be expected by chance would have been $(1/4)\ \binom{n}{3}$, and the variance of the distribution would have been $(3/16)\ \binom{n}{3}$ (Kendall, 1955, p. 157); in this case the expected value was 242 and the standard deviation was 13.5. The distribution for this number of triads was approximately normal, and it can be seen that there was only one student within three standard deviations of the chance expectation.

The pair-comparison data for the total group is presented in Table 1 (1966 data), each cell containing the proportion of subjects who judged

---

[4] This distribution is for the 311 subjects who responded to all 171 pair comparisons.

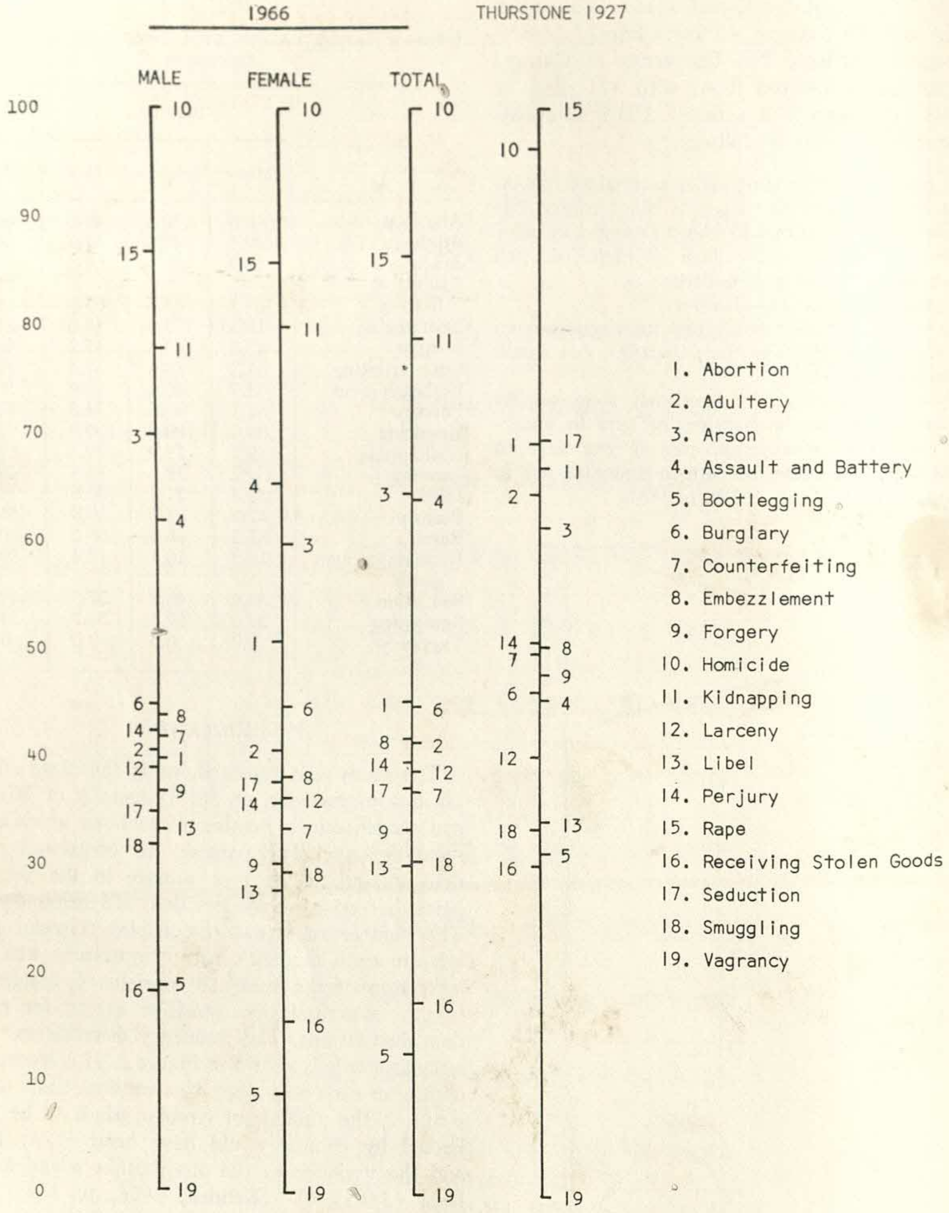


FIG. 2. Scales for all offenses.

the row offense to be more serious than the column offense. The rows and columns are in alphabetical order. Thurstone's data are available in his article and are not reproduced here.

The 1966 data could be analyzed separately for males and females as well as for the total group, so altogether four scales were constructed,[5] utilizing Case V of the law of comparative judgment.

Table 2 gives the scale values that were obtained from each set of data with the scales arbitrarily equated at both extremes (0 and 100).

Figure 2 illustrates the four scales—the numerical code is the alphabetical order.

Thurstone classified the offenses into three groups: sex offenses, injury to the person, and property offenses. For clarity of comparison, Figures 3, 4, and 5 display the relative positions of

---

[5] Thurstone's data were reanalyzed in order to insure that exactly the same procedure was used on all data. There were no differences between his results and the reanalysis reported here.

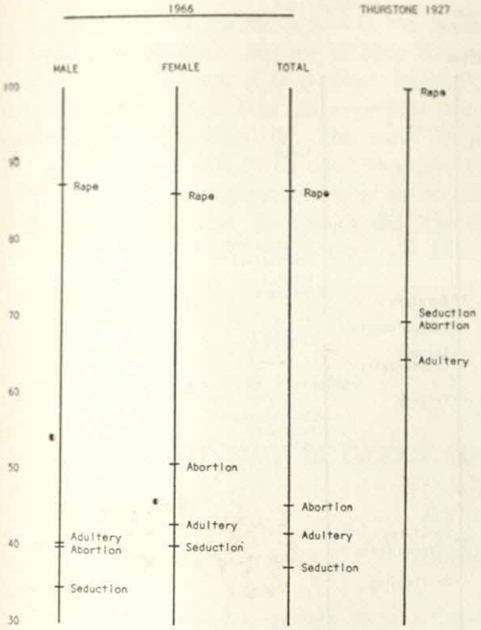


FIG. 3. Scales for sex offenses.

the offenses in each group, respectively, for the different sets of data.

INTERPRETATION

In interpreting these results, it should be emphasized that it was not the college students' legal knowledge which was being tested, but their attitudes toward the seriousness of these offenses, which they may have known only through the daily newspaper. Also, the samples of students were only superficially matched. The students who went to the University of Chicago in 1926 probably represented a narrower segment of the society, socially and economically, than do college students today, when a larger portion of the population is attending college. It may also be conjectured that there were differences in the caliber of the students in the two samples, but it is not obvious what effect, if any, such differences would have had on the results.

It should also be kept in mind that the origin and the unit of measurement for each scale was arbitrary, so that all changes were relative. For example, it cannot be said that Crime X is re-

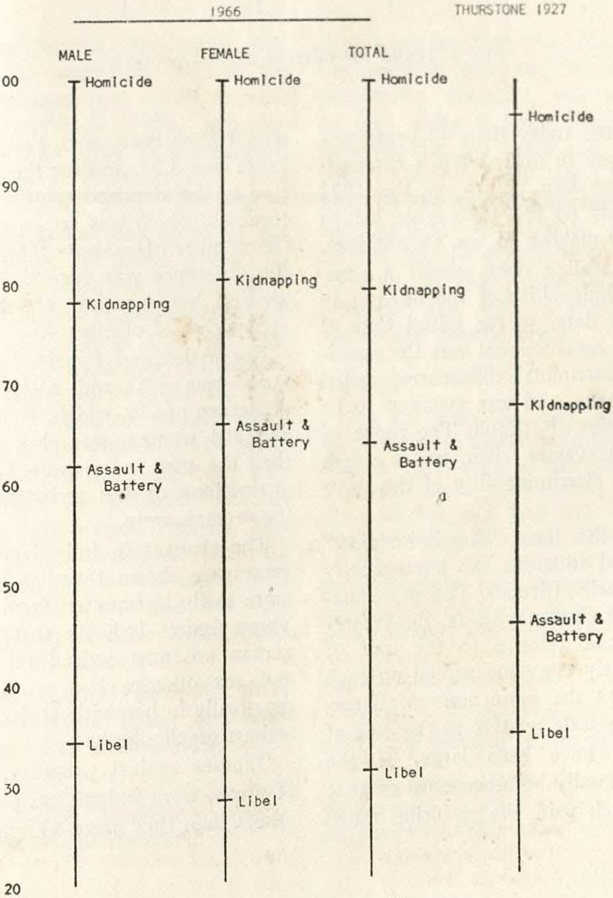


FIG. 4. Scales for injury to the person.

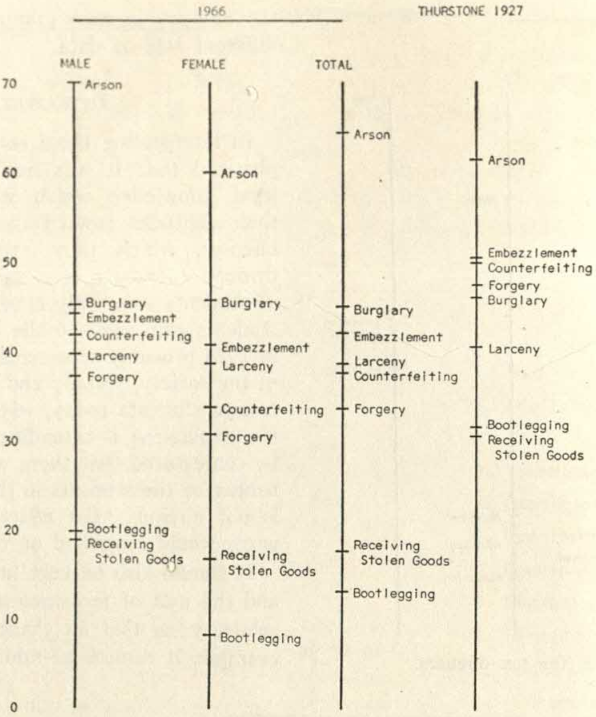


FIG. 5. Scales for offenses against property.

garded as more serious today than 40 years ago. There is no valid way to make such a comparison without equating two points on the 1927 scale with two points on the 1966 scale, which can only be done by making ad hoc assumptions.

This method of scaling does permit a comparison of the discriminability of the offenses in the earlier and later data. In the initial Case V solution, the unit of measurement was the standard error of the discriminal differences, called the comparatal variance, and was assumed to be a constant for all pairs of stimuli. The range of the original raw-scale values, then, was a rough index of the relative discriminability of the stimuli.

In this context, the term "discriminability" should be interpreted to mean the homogeneity of the subjects in their attitudes. The more uniformity or agreement among subjects, the greater the range of the scale relative to the unit of measurement. If nearly everyone agreed on each pair as to which was the more serious offense, then the range of the scale relative to the unit of measurement would have been large. If the group had been maximally heterogeneous so that it split 50/50 on each pair, the resulting range would have been zero. For Thurstone's data the range was 3.28, and for the 1966 data, 3.02, relative to the standard error of the discriminal differences, which was set equal to *one*. In view of the number of subjects these data were based on, the difference was very significant and indicated greater homogeneity of attitudes toward the seriousness of offenses 40 years ago than now.

For males and females in the 1966 data, the range was 3.20 and 2.94, respectively; also a difference of .26, which, being based on 165 men and 204 women, was also significant, suggesting that the men were more homogeneous in their evaluations of the seriousness of offenses than the women were.

The changes in judged seriousness for all offenses are shown together in Figure 2, but are more easily interpreted from Figures 3, 4, and 5. These figures indicate that offenses against the person are now judged relatively more serious and sex offenses less so. This is shown most drastically in homicide replacing rape as the most serious of all offenses.

Offenses against property, with the exception of arson, were judged less serious relative to the others than they were 40 years ago.

The principal sex difference was in the judged seriousness of abortion, women judging it more serious than men; and it is perhaps interesting that the men judged rape as somewhat more serious than the women did. The other major sex difference was that, with one exception, the men judged offenses against property to be relatively more serious than the women did. The exception was burglary, for which there was almost no difference.

REFERENCES

KENDALL, M. G. *Rank correlation methods.* (2nd ed.) New York: Hafner, 1955.

THURSTONE, L. L. A law of comparative judgment. *Psychological Review,* 1927, **34,** 273–286. (a)

THURSTONE, L. L. Method of paired comparisons for social values. *Journal of Abnormal and Social Psychology,* 1927, **21,** 384–400. (b)

THURSTONE, L. L. *The measurement of values.* Chicago: University of Chicago Press, 1959.

(Received April 25, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 91–92

RELATIONSHIP BETWEEN LOCUS OF CONTROL AND ANXIETY[1]

DAVID WATSON

University of Toronto

Correlations between locus of control, manifest anxiety, and test anxiety (Alpert-Haber Achievement Anxiety Test) were computed for a large sample—N = 648. There were significant relationships between all variables.

Assuming that those persons who cannot complete organized response sequences will, in the event that no other response is available, become anxious, Mandler and Watson (1966) hypothesized that individuals who score in the external direction on the Locus of Control (LC) scale (Rotter, Seeman, & Liverant, 1962) will tend to be more anxious than those who score in the internal direction, because the latter group will more often appraise the world as one in which they can complete organized response sequences. Indeed, Rotter (1966) reports that one researcher—Ware—found a significant correlation ($r = 24$, $N = 111$, $p < .05$) between the LC scale and the Taylor (1953) Manifest Anxiety scale (*MA* scale), but that Efran with $N = 114$ found no significant relationship between the two. Butterfield (1964) correlated the LC scale with the Alpert-Haber (1960) Achievement Anxiety Test (AAT) and reported significant relationships with both parts of the AAT; $r = .61$ ($p < .01$) with debilitating anxiety (AAT Deb) and $r = -.82$ ($p < .01$) with facilitating anxiety (AAT Fac). However, Butterfield's N was only 47.

Mindful of Cohen's (1962) plea for more consideration for the statistical power of designs and analyses of data, and of Milholland's (1964) demonstration that this consideration is indeed important in analyzing the relationships between personality variables, this study was designed to test, using a large sample, the relationship between the LC scale and the *MA* scale and AAT. Such a study, too, would shed further light on the relationship between the *MA* scale and the AAT—reported by Alpert and Haber to be $r = .38$, ($p < .05$), but with an N of only 40—and on the relationship between the two parts of the AAT, discussed by Alpert and Haber and several others.

Method

The LC scale, the AAT, and 26 items[2] of the MA scale were administered to all students in introductory psychology at the University of Toronto. The AAT items were scored on a 10-point scale, the AAT Fac and AAT Deb subscales being scored separately. The MA scale items were answered true or false, and the LC scale items were presented in forced-choice pairs. Pearson product-moment correlations between the three scales were calculated by a computer for 506 females and 142 males.

Results and Discussion

Tables 1, 2, and 3 display the correlations for males and females separately and for sexes com-

---

[1] This study was supported by Grant No. 67 from the Ontario Mental Health Foundation. Evelyn Baumal aided in the analysis of the data.

[2] The revised (short) form of the *MA* scale was used, except for the two items numbered 25 and 44 in Taylor's (1953) report.

TABLE 1

CORRELATION COEFFICIENTS FOR MALES

	MA scale	AAT Deb	AAT Fac
LC scale	.38**	.25**	−.15*
MA scale		.52**	−.25**
AAT Deb			−.41**

Note.—$N = 142$.
* $p < .06$.
** $p < .01$.

TABLE 3

CORRELATION COEFFICIENTS FOR SEXES COMBINED

	MA scale	AAT Deb	AAT Fac
LC scale	.36**	.25**	−.08*
MA scale		.46**	−.19**
AAT Deb			−.44**

Note.—$N = 648$.
* $p < .05$.
** $p < .01$.

bined. Almost all the relationships are significant, though some are rather uninterestingly small.

The significant correlation between the *MA* scale and the LC scale indicates that the more external is one's appraised locus of control, the more anxiety one reports. The coefficients are larger than those observed by Ware, but are probably within the fiducial limits of his coefficients. Efran, with a sample size nearly identical to Ware's, could have easily obtained a nonsignificant relationship and yet remained within the fiducial limits. The confidence belt for the present correlations, of course, is much narrower because the *N* is so large.

The significant correlations between AAT Deb and the LC scale also indicate that the more external is one's locus of control, the more anxiety reported. The coefficients here are much smaller than those reported by Butterfield. Since the *N*s used in the present analyses are much larger than Butterfield's, we feel that the coefficients are a more valid reflection of the relationship.

The relationship between locus of control and anxiety seems well supported by these correlations, and these results lend credence to Mandler and Watson's hypothesis that actual or appraised lack of control produces anxiety.

The relationship between *MA* scale and AAT Deb is larger than that reported by Alpert and Haber, but the correlation between *MA* scale and AAT Fac is smaller. The *MA* scale appears to be more closely related to "debilitating" than to "facilitating" anxiety. The same is true of the relationship between the LC scale and the AAT. The concept of "facilitating" anxiety can, in any case, be confusing, as anxiety is often understood to mean a nonfacilitative level of motivation. Thus one would not necessarily expect the Fac scale to correlate highly with scales which probably measure nonfacilitative motivational levels. The relationship here between AAT Deb and AAT Fac is very similar to that reported by Alpert and Haber.

TABLE 2

CORRELATION COEFFICIENTS FOR FEMALES

	MA scale	AAT Deb	AAT Fac
LC scale	.35**	.26**	−.06
MA scale		.44**	−.17**
AAT Deb			−.44**

Note.—$N = 506$.
** $p < .01$.

For all of these relationships, the much larger *N* of this study gives greater credence to the present results.

REFERENCES

ALPERT, R., & HABER, R. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology,* 1960, **61**, 207–215.

BUTTERFIELD, E. Locus of control, test anxiety, reactions to frustration and achievement attitudes. *Journal of Personality,* 1964, **32**, 355–370.

COHEN, J. The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology,* 1962, **65**, 145–153.

MANDLER, G., & WATSON, D. L. Anxiety and the interruption of behavior. In C. Spielberger (Ed.), *Anxiety and behavior.* New York: Academic Press, 1966. Pp. 263–288.

MILHOLLAND, J. Note on the further validation of the Alpert-Haber Achievement Anxiety Test. *Journal of Abnormal and Social Psychology,* 1964, **69**, 236.

ROTTER, J. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs,* 1966, **80**(1, Whole No. 609).

ROTTER, J., SEEMAN, M., & LIVERANT, S. Internal versus external control of reinforcement: A major variable in behavior theory. In N. Washburne (Ed.), *Decisions, values and groups.* Vol. 2. New York: Pergamon Press, 1962. Pp. 473–516.

TAYLOR, J. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology,* 1953, **48**, 285–290.

(Received May 9, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 93–98

CENSURE OF THE MODEL IN THE CONTAGION OF AGGRESSION[1]

LADD WHEELER AND SEWARD SMITH[2]

Naval Medical Research Institute, Bethesda, Maryland

119 Ss participated in a 6-condition experiment involving a "discussion" between S and 3 tape-recorded confederates. The dependent variable was S's verbal aggression toward Confederate A—as a function of the deviancy of A's opinions, the modeling of aggression toward A by Confederate B, and the censuring of B for his aggression. The results were that restraints against aggression were reduced by B's modeling of aggression and were reinstated by E censure of B. B's self-censure or censure of B by Confederate C did not reinstate restraints. Postexperimental liking for B was least when he had unjustifiably aggressed and when he had been censured for justifiable aggression by E or by Confederate C. Liking for B was greatest when he had refrained from justifiable aggression and when he had censured himself for justifiable aggression. Reinstatement of restraints by E censure produced high scores on a depression scale.

Behavioral contagion has recently been defined by the following set of conditions: (*a*) A person is instigated to perform an act; (*b*) he does not perform the act; (*c*) he observes someone else (the model) perform the act; and (*d*) he then performs the act (Wheeler, 1966). Implicit in this definition is the assumption that observing someone else perform the act reduces the observer's restraints. However, according to the theory of behavioral contagion, the observer's restraints are not always reduced; whether restraint reduction takes place depends in part upon the observable consequences accruing to the model for his behavior. Or stated more precisely, restraints momentarily reduced by observation of the model may be immediately reinstated if the model is punished.

Several studies have shown the importance of response consequences to the model. Bandura, Ross, and Ross (1963) exposed children to a televised aggressive model who was either rewarded or punished for his aggression. Subsequently, subjects who had observed the rewarded model were more aggressive than those who had observed the punished model. Walters, Leat, and Mezei (1963) forbade children to play with a set of toys and then exposed the children to a film showing a child either rewarded or punished by his mother for playing with the toys. Children who observed the rewarded model played with toys more frequently. Walters and Parke (1964) replicated these results and further demonstrated that the children in the two treatment groups played equally with the toys once they were no longer forbidden to do so. Thus, response consequences to the model affected performance rather than learning.

In each of the studies cited above, the subjects were young children; there is no direct evidence that response consequences to the model influence an adult observer. Furthermore, in each of the studies, the model was presented in a film medium, which, especially for children, may be unusually persuasive. Finally, the studies were not concerned with the characteristics of the punishing agent.

A different experimental approach has been taken in recent research on the contagion of aggression (Wheeler & Caggiula, 1966). In these experiments subjects were adult males who interacted directly with the model and with the target of aggression. A similar technique was used in the present research to determine (*a*) whether response consequences (verbal censure) to the model affect the contagion of aggression among adults, and (*b*) whether there is a differential effect of censure delivered by the experimenter, a peer, or the model himself.

Several predictions were tested in the present experiment. It was predicted that observing a noncensured model would effectively reduce restraints against aggression and that censure of the model by the powerful and high-status experimenter would effectively reinstate these restraints. It was predicted that peer censure would not reinstate restraints because the peer lacked any formal power over the subject, was of no higher status than the subject, and, being a stranger to the subject, also lacked normative influence. Self-censure, or censure of the model by himself, was predicted to fall at an intermediate position between experimenter censure and peer censure. The basis of this latter prediction is that, while the model also lacked power, status, or normative influence with regard to the subject, he indicated by the very act of aggression some similarity to the subject in affective response. To the extent that the subject generalizes from this one instance of similarity in affective response, he should anticipate the same regret following aggression as that expressed by the model.

---

[1] From Bureau of Medicine and Surgery, Navy Department, Research Task MF022.01.03-1002. The opinions and statements contained herein are the private ones of the writers and are not to be construed as official or as reflecting the views of the Navy Department or the Naval Service at large.

[2] We thank Ira Donenfeld for his highly competent conduct of this experiment and many of the analyses. We also note with appreciation the contributions of Peter M. Edmondo for instrumentation and analysis assistance, and of Hank Davis, Morris D. Stanton, and George J. Tresansky, who portrayed the confederates.

In addition to overt verbal aggression, data were collected on the subject's affect and his perception of the other group members. It was felt important to determine the relationship between the effectiveness of censure and the residual affect and interpersonal attitudes, because what may be an effective group-management technique by one criterion may be undesirable by another.

METHOD

Overview

The subject and three tape-recorded confederates engaged in a "discussion." On each of two topics, Confederate A stated his opinion, Confederate B evaluated A's opinion, Confederate C evaluated A's opinion, and the subject evaluated A's opinion. A's opinions were either socially desirable or undesirable, and B's evaluations were either aggressive or nonaggressive. When B aggressed, he immediately censured himself for having aggressed, was censured by the experimenter, was censured by C, or was not censured at all. The subject's aggression toward A was the major dependent variable.

Subjects

Subjects were enlisted men from the Bainbridge, Maryland, Naval Training Station, with slightly above average IQ, and an average age of 19.3.

Procedure

The subject was taken to one of several adjoining soundproof cubicles and told that he would be talking via an intercom system with three other enlisted men whom he did not know and would not meet. He was told that full instructions regarding the discussion would be given by Commander Wheeler over the intercom system. To protect the subject's identity, he was to be known simply as Man D. Taped instructions informed the subject that each of the four men was to state his opinions on two separate topics and that each opinion would be evaluated by the other three men. In actuality, the experiment was terminated after Man A had stated his two opinions and had been evaluated by the other three men. The two topics on which A stated his opinions were *parents* and *religion,* in that order for half the subjects and in the opposite order for the remaining subjects. Forty-five seconds were allowed for each opinion and 30 seconds for each evaluation. The subject's evaluation was accomplished by selecting the phrase from the "Reaction to Opinion List" that best described how he felt toward A's opinion and then stating why he felt that way. The Reaction to Opinion List was in actuality the Berkun Hostility Scale (Berkun, Burdick, & Woodring, in press).[3]

When the subject selected a phrase, which ranged from "Delighted with" (0) to "Raging mad at" (97), the associated scale value was used as an index of overt aggression by the subject against Man A. After A's two opinions and the corresponding evaluations, the subject was requested to fill out questionnaires designed to measure his general mood and his attitude toward the other men.

Conditions[4]

No instigation. In this condition, and in this condition only, A expressed socially desirable, moderate, and rational opinions, thus creating no instigation toward aggression in the subject. Nevertheless, B aggressed quite strongly against A during the evaluation period. On both topics, he chose the phrase "fed up with" and then proceeded to curse A in a rather vile and violent manner. He provided no rational objections to A's opinion. C, in turn, chose the phrase "indifferent toward" and stammered around, communicating that he just did not know what else to say.

No aggression. A expressed extreme and socially undesirable opinions. On the topic *parents,* for example, he stated that there was no reason to respect parents at all because the only reason they had children was "for their own fun in bed." On the topic *religion,* he stated that all religions other than his own "should be made illegal." In the evaluation period, B chose the phrase "displeased with" and followed this by very calmly stating that he believed that he held a different opinion from A on this topic. C again chose the phrase "indifferent toward" and could not find anything else to say about it.

---

[3] One-hundred twenty-six Army basic trainees placed 90 phrases in 1 of 11 categories of increasing hostility. Sixteen phrases were chosen on the basis of having medians spaced as evenly as possible among the 11 categories and having small interquartile range values.

[4] The reader interested in complex designs may be interested in the fact that this is an incomplete nested hypercube. W. W. Haythorn brought this to our attention.

TABLE 1

MEAN HOSTILITY SCALE SCORES BY CONDITION

N	Condition	$\bar{X}$	Nearest scale phrase
(18)	No instigation	36.3_a	Did not mind
(20)	*E* censure	53.8_b	Displeased with
(15)	No aggression	54.8_{bc}	Displeased with
(20)	Self-censure	59.5_{cd}	Offended at
(24)	Peer censure	61.1_d	Offended at
(22)	No censure	63.0_d	Cross with

Note.—Conditions having no subscripts in common differ by t test with $p < .05$ or better.

No censure. A expressed his extreme opinions as in the no-aggression condition. B aggressed as in the no-instigation condition, and C reacted in his usual noncommittal fashion.

Experimenter censure. This condition was identical to the no-censure condition except that at the end of B's evaluation period the experimenter, whom the subjects believed to be Commander Wheeler, broke in and publicly censured B for his aggression. His statement on one topic was, "I'm sorry you said that, B. It's no good flying off the handle like that." On the other topic, he said, "You lost your temper, B. I'm sorry you did, because everybody has a right to his own opinion."

Peer censure. This condition was also like the no-censure condition except that C, after having given his usual noncommittal evaluation of A's opinion, censured B for aggression, using the same phrases that the experimenter used in the experimenter-censure condition.

Self-censure. Once again, this condition was identical to the no-censure condition, with one exception. After having aggressed and paused for a few seconds, as if in thought, B censured himself for his aggression using the same phrases used in the experimenter-censure and peer-censure conditions. Only referents were changed—for example, "I lost my temper, A"; instead of "You lost your temper, B."

RESULTS

Overt Aggression

The measure of overt aggression was the scale value of the word from the Berkun Hostility Scale which was openly communicated by the subject to Man A. By t test, no condition differed significantly ($p < .05$)[5] from the first to the second topic. Accordingly, the scale values from the two topics were combined to produce the mean aggression shown in Table 1.

It is apparent from Table 1 that the data are in general accord with the predictions. Compared to subjects in the unqualified contagion condition (no censure), subjects in the control groups (no instigation and no aggression) exhibited relatively little aggression. The predicted and obtained ordering of the contagion conditions, from most to least aggression, was (*a*) no censure, (*b*) peer censure, (*c*) self-censure, and (*d*) experimenter censure. The a priori probability of precisely this ordering was less than .042, this being one of 24 possible orderings. The mean differences, however, are small relative to the variance, and the only censure group to differ significantly from the no-censure condition was experimenter censure. In addition, the experimenter-censure group differed significantly from the other two censure conditions. Further, experimenter censure produced even slightly less aggression than the no-aggression condition, indicating that restraints reduced by the aggressive model were completely reinstated by experimenter censure of the model.

The data are ambiguous only with regard to self-censure. On the one hand it can be argued that self-censure *was not* effective, in that this condition did not differ significantly from the no-censure condition. On the other hand, there is equally good evidence that self-censure *was* effective in reinstating restraints, because the condition did not differ significantly from the no-aggression condition. In an attempt to reduce this ambiguity, a different type of analysis was performed.

Berkun Hostility Scale scores were averaged over the two topics, and an average score of 56 or greater was termed an aggressive response. The number 56 was chosen because it is the scale value of "displeased with," the first word on the scale indicating aggression (the next lowest scale value is associated with "indifferent toward").

Using this definition of aggression, 40% of the subjects in the no-aggression condition were aggressive, with corresponding values in the self-censure and no-censure conditions of 75% and 81.8%. By exact test, the no-aggression and self-censure conditions differed from one another with $p = .08$, which we will accept as evidence of a real difference. Thus, self-censure was not effective in reinstating restraints against aggression.

Again using the simple dichotomy of aggression versus nonaggression it was possible to test, as Wheeler and Caggiula (1966) did, the restraint reduction hypothesis that (*a*) observation of an aggressive model, combined with (*b*) instigation to aggression, produces a higher frequency of aggression than would be predicted by a simple additive combination of *a* and *b*. In the no-instigation group, 5.9% of the subjects aggressed, while 40.0% did in the no-aggression condition, giving an additive prediction of 45.9% in the no-censure condition. Comparing the predicted 45.9% with the observed 81.8% produces a goodness of fit chi-square of 11.42 ($df = 1$, $p < .001$).[6]

---

[5] All significance levels are reported as two-tailed.

Liking for Other Group Members

After the two topics had been discussed, but before the end of the experiment had been announced, subjects were asked to rate Men A, B, and C on six evaluative 7-point scales, ranging from −3 to +3. Scores from the six scales were combined to yield the condition by ratee means presented in Table 2.[7]

The most deviant means in Table 2 are from the no-instigation condition, and this is precisely as it should be; in this condition, Man A voiced reasonable opinions and yet was attacked by Man B. Thus, the former was liked and the latter disliked, providing additional confirmation of the success of the manipulation.

The data from the remaining five treatment conditions (excluding the no-instigation condition) are, however, of psychological interest. For these five conditions, data for each of the three men were separately subjected to a one-way analysis of variance. Differences between groups did not approach significance with regard to Man A or Man C. Lack of a difference with regard to Man A confirmed the assumption of equivalent instigation in the several anger conditions. With regard to Man B, the conditions differed at $p < .10$. The following significant differences between groups in liking of Man B were indicated by t tests: the no-aggression group mean was higher than the experimenter-censure ($p < .01$) and peer-censure means ($p < .05$), and the self-censure mean was higher than the experimenter-censure mean ($p < .05$) and the peer-censure mean ($p < .10$). Thus, it appears that if popularity is a goal, one should either not violate the social prohibition against aggression or that one should immediately apologize for doing so—particularly, one should not wait to be criticized by others for his violation.

TABLE 2

LIKING FOR OTHER GROUP MEMBERS

Condition	Man A	Man B	Man C
No instigation	8.0	−7.7	3.0
E censure	−8.0	2.1	4.5
No aggression	−9.7	7.5	3.1
Self-censure	−10.0	6.8	4.6
Peer censure	−10.3	2.7	3.7
No censure	−8.9	5.0	3.7

Note.—A positive number indicates liking.

Mood Changes during the "Discussion"

Mood was also assessed after the discussion, using the Primary Affect Scale (PAS).[8] The scale consists of 10 series of words or phrases (five or six choices each). From each series the subject is asked to select the word or phrase that most and second most applies to his feelings during a given time. From the subject's answers are derived scale values for amount of rated happiness, anger, fear, depression, and arousal (alertness). In this experiment subjects filled out the PAS twice, once to indicate how they felt usually during a discussion with other sailors, and again to reveal how they felt during the "discussion" which had just ended.

The change scores ("during" minus "usually") were the data of greatest interest.

Table 3 presents the mean change scores for each mood factor by treatment condition. Generally, the no-instigation subjects experienced the least change in mood, as would be expected. In order to assess overall mood shift for each treat-

TABLE 3

CHANGE SCORES ("DURING" MINUS "USUALLY") FOR THE FIVE MOOD FACTORS AS A FUNCTION OF TREATMENT CONDITION

Mood factor	No instigation	No aggress	*E* censure	Self-censure	Peer censure	No censure	All groups
Happiness	$.0_a$	$-.7_{abc}$	$-.5_{ab}$	-1.8_{bc}	-2.2_c	-1.7_{bc}	−1.1
Anger	1.2_{ab}	3.2_c	3.0_{abc}	$.8_a$	2.2_{abc}	2.4_{abc}	2.1
Fear	$.2_a$	1.6_b	$.5_{ab}$	1.1_{ab}	1.0_{ab}	$.9_{ab}$	.9
Depression	$-.1_a$	1.2_b	2.9_c	$.8_{ab}$	$.6_{ab}$	$.9_{ab}$	1.0
Alertness	$.1_a$	$.1_a$	$-.3_a$	$-.5_a$	$-.5_a$	$.1_a$	−.2
(A + F + D − H)	1.3	6.7	6.9	4.5	6.0	5.9	

Note.—For each mood factor, conditions having no subscripts in common differ by t test with $p < .05$ or better.

---

[6] $[.059 + .40 - (.059)\ (.40)]\ 100 = 43.5\%$ is technically the additive prediction. The more conservative 45.9% was used for simplicity.

[7] The positive ends of the scales were "intelligent," "likable," "brave," "honest," "minds his own business," and "fair."

[8] The PAS was developed by Thomas I. Myers, Eugene Johnson, III, and Seward Smith, Naval Medical Research Institute, in order to assess affective states in each of five a priori categories—happiness, anger, fear, depression, and arousal.

ment group, a total change score was derived (anger + fear + depression − happiness: alertness was not included in the total change score because there was no reason to believe it relevant). Using this score, also shown in Table 3 for each condition, the no-instigation group had a significantly ($p < .02$ or better) less negative mood shift than did all other treatment groups. No other groups differed from one another.

Comparisons (by means of *t* tests) between treatment groups for the various mood factors were made. Ignoring the several points of difference between the no-instigation group and the other groups, relatively few significant differences remain. Only on the mood factor of depression were there several differences among conditions. These differences all involved the experimenter-censure group, which was significantly ($p < .05$ or better) more depressed than *any* other condition.

DISCUSSION

The major concern of this research was with censure of an aggressive model in a small group. We wished to know if censure of the model could reduce or eliminate subsequent aggression within the group and whether one agent of censure was more effective than another. Further, it was desirable to understand other, less overt, effects of model censure—such as mood and interpersonal attitudes.

That experimenter censure should have been so effective is not surprising in view of the power and status invested in the experimenter. Not only did he have the usual scientific status, but he was also alleged to be a high-ranking naval officer (compared with the enlisted status of the subjects). Had the experimenter been a graduate assistant, results might have been quite different.

The cards were stacked against the reinstatement of restraints by peer censure. Subjects had not met the peer and did not expect to meet him. The personality he conveyed was that of a not-too-bright, somewhat withdrawn individual who did not even voice disagreement with the extreme opinions of the target. We would expect peer censure to be far more effective if there were stronger affectional bonds between the subject and the peer, if the subject expected future interaction with the peer, or if the peer were of high status.

The effect of self-censure was not too clear. In order to conclude that self-censure was not effective it was necessary to conduct a separate and somewhat ad hoc analysis. In general we would expect self-censure to be effective to the extent that the censure appeared to be sincere and deeply felt. Our model did appear sincere in censuring himself for aggression, but perhaps did not appear strongly *regretful.*

There is one important point which applies to each of the censure conditions; it is that censure on the first topic did not affect the model's behavior on the second topic—he aggressed at the same level of intensity on both. Had the model not aggressed on the second topic after being censured for aggressing on the first, the subject might have attached greater meaning to the censure. Also, this heavier weighting of *effective* censure might be more pronounced for one type of censure than for another. Our guess is that self-censure in particular would increase in effectiveness as a result of being shown to influence the model's subsequent behavior.

With regard to subjects' postexperimental liking for the other members of the group, the major result was that subjects in the experimenter-censure and peer-censure conditions were not as strongly attracted to the model as were subjects in the no-aggression and self-censure groups. These differences are not consistent with Wheeler's (1966) prediction that models who successfully reduce restraints will be liked. The peer-censured model certainly reduced restraints against aggression, but was not liked as well as the calm model of the no-aggression condition, who did not reduce restraints. Perhaps increased liking for a model occurs only when he is rewarded for his behavior, thus removing the observer's anxiety. In the current experiment, liking of the model appeared to be a function of whether or not he was externally censured. Thus it may be that while peer censure was ineffective in reinstating controlling restraints against aggression, it was effective in producing some degree of anxiety about it. Experimenter censure was effective both in reinstating controlling restraints and producing anxiety. On the other hand, it may be that aggressive models are usually disliked regardless of the consequences accruing to them. Hartmann (1965) found that subjects who observed a filmed model quite successfully beat a victim liked the model less than they did the victim—even though these subjects were clearly influenced by the model to aggress toward a peer.

In conclusion, the mood or affect data have interesting implications. Mood strongly differentiated the no-instigation treatment from all other groups, who showed pronounced negative mood shifts. Generally there were no differences among these latter groups. The exception was the factor of depression, which was greater for the experimenter-censure group than for any other. Only in this condition were restraints against aggression reduced and then immediately reinstated before the subject had an opportunity to emit any behavior. If one has the practical aim of minimizing aggression within a group, he must remain aware of the fact that authoritarian censure of the person who expresses the dominant group feeling may result in a feeling of depression even more damaging than the free expression of aggression.

REFERENCES

BANDURA, A., ROSS, D., & ROSS, S. A. Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology,* 1963, **66,** 601–607.

BERKUN, M. M., BURDICK, H. A., & WOODRING, T. M. A check list for feelings of hostility. *Psychological Reports,* in press.

HARTMANN, D. P. The influence of symbolically modeled instrumental aggressive and pain cues on the disinhibition of aggressive behavior. Unpublished doctoral dissertation, Stanford University, 1965.

WALTERS, R. H., LEAT, M., & MEZEI, L. Response inhibition and disinhibition through empathetic learning. *Canadian Journal of Psychology,* 1963, **16,** 235–243.

WALTERS, R. H., & PARKE, R. D. Influence of response consequences to a social model on resistance to deviation. *Journal of Experimental Child Psychology,* 1964, **1,** 269–280.

WHEELER, L. Toward a theory of behavioral contagion. *Psychological Review,* 1966, **73,** 179–192.

WHEELER, L., & CAGGIULA, A. R. The contagion of aggression. *Journal of Experimental Social Psychology,* 1966, **2,** 1–10.

(Received May 9, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 98–101

EFFECT ON OPINION CHANGE OF HOW DESIRABLE THE COMMUNICATION IS TO THE AUDIENCE THE COMMUNICATOR ADDRESSED[1]

JUDSON MILLS AND JERALD M. JELLISON
University of Missouri

An experiment tested the hypothesis that when the audience feels the communicator thinks the communication will be desirable to the audience that is addressed they will be less persuaded than when they feel the communicator thinks the communication will be undesirable to the audience addressed. College students read a speech favoring tripling truck license fees. Some were told it was delivered to railway men (desirable condition), others, that it was delivered to truck drivers (undesirable condition). Agreement with the communicator was lower in the desirable condition than the undesirable condition. The results confirm the hypothesis, which provides an explanation for previous findings that overheard communications were more persuasive than regular communications only when the conclusions were desirable.

In their experiment on the effectiveness of overheard communications, Walster and Festinger (1962) found that a communication which subjects thought they had overheard was more effective than a regular communication only when the topic of the communication was "personally involving." For example, when the communication argued that husbands should spend more time at home, married women were more influenced if they thought they had overheard the communication, but there was no difference for single women. In a rather well-controlled design, Walster and Festinger showed that the same subjects who were more affected by an overheard communication when the issue was personally involving were not more affected by an overheard communication when the issue did not involve them. The involving communications which they used all argued in favor of a position that was desirable, something the subjects would want to believe. Walster and Festinger noted that this creates a problem in interpreting their results. They commented,

> There is, of course, still the interesting question, to which we have no answer, as to why, when the involved subjects hear a communi-

---

[1] This study was supported by a grant from the National Science Foundation.

cation *which advocates a position they would like to accept,* they still tend to impute ulterior motives to the speaker when this is at all possible [p. 401].

In a follow-up study, Brock and Becker (1965) found some evidence that an overheard communication was more effective than a regular communication when the topic was involving and the conclusion desirable, but not when it was involving and undesirable or uninvolving.

It is possible that the subjects in these studies who received the involving communications with desirable conclusions and thought the communicator was aware they were listening (the regular communication) also assumed that the communicator was aware that the communication would be desirable to them. When the subjects thought they had overheard the communicator, there is no reason to expect that they assumed that the communicator was aware that the communication would be desirable to them. It is reasonable to suppose that when the audience feels that the communicator thinks the communication would be desirable to the audience they might tend to doubt his sincerity, to suspect that he is just saying something to please them, and thus may be less influenced. If, on the other hand, the audience feels that the communicator thinks the communication would be undesirable to the audience, something they would not like to believe, they would not tend to suspect his motives and might even be more influenced.

An experiment was designed to test the hypothesis that when the audience feels that the communicator thinks the communication will be desirable to the audience that is addressed they will be less persuaded than when they feel the communicator thinks the communication will be undesirable to the audience addressed.

METHOD

College students read a speech favoring increasing tractor-trailer license fees under two experimental conditions. In the desirable condition, the speech was described as having been delivered before a local union of railway men. In the undesirable condition it was supposedly delivered before a local union of long-haul truck drivers. Agreement with the communicator's position was measured after the subjects had rated his personality characteristics and completed a test of memory for what was said in the speech.

The subjects were students in introductory psychology who participated in the experiment as part of a requirement in their course. They were run in groups composed entirely of the same sex. Within each group, subjects were randomly assigned to the two experimental conditions, with as near to an equal number as possible in each condition. There were a total of 59 male and 19 female subjects in the desirable condition and 60 male and 18 female subjects in the undesirable condition.

When the subjects in a particular group had assembled, the experimenter told them that the study concerned impression formation. Each of them would read a speech and then rate the personality of the man who delivered it. Booklets were distributed containing a fictitious speech on a proposed bill to triple state license fees for truck tractors and trailers. The introduction to the speech stated that at the time it was given the speaker was a candidate in the primary election for a seat in the Missouri legislature from the St. Louis area. In the desirable condition, the introduction added that the speech was given at a meeting of a local union of railway men in St. Louis. In the undesirable condition, it stated that the speech was given at a meeting of a local union of long-haul truck drivers in St. Louis.

In the five-page speech, which was identical in both conditions, the speaker took a strong stand in favor of the proposed bill to triple the license fees of tractor trailers. The speaker's main arguments were that truckers were not presently paying their fair share of the costs of highway construction and maintenance, and that the bill would increase competition between the railroads and trailer trucks, which would strengthen national defense.

When the subjects had finished reading the speech booklets, they were collected, and each subject was given a rating form with the same code number as their speech booklet. The subjects were asked to indicate how well various characteristics applied to the speaker by circling a number from 0 (extremely inappropriate) to 20 (extremely appropriate) for each characteristic. The characteristics were amiable, cynical, honest, impartial, likable, obliging, opinionated, opportunistic, sincere, stupid, tactful, unaccommodating, and unconventional. The subjects were told that their ratings would be kept completely confidential and were not asked to put their names on the rating forms.

After the subjects had completed the rating forms, the experimenter "confessed" that there was more to the study than he had mentioned at the beginning. He told them that another purpose was to study the relationship between impressions of a person and memory for what the person had said. The experimenter explained to the subjects that he was now going to pass out a memory test to determine how well they could recall exactly what the speaker said in the speech. He explained that he could not reveal this earlier because if he had they might have made a special effort to memorize the speech. The subjects were warned that the memory test was difficult.

The rating forms were collected, and each subject was given a booklet for the "memory test" with the same code number as their speech booklet and rating form. Subjects were not asked to put their names on the memory-test booklet. The first section of the booklet contained 23 multiple-choice-type questions with three alternatives each. The subjects were instructed to choose the word or phrase that completed each statement in the way which was closest to what the speaker said. The second item, which was not included in the calculation of the recall scores, was:

> The speech was given at a meeting of a union of ______.
>
> a. truck drivers
> b. railway men
> c. airline employees

Seven male and three female subjects in the desirable condition failed to choose alternative b for this item, and one male and four female subjects in the undesirable condition failed to choose alternative a. These subjects were excluded from the analysis of results.

After all the memory-test booklets were distributed, the experimenter interrupted the subjects and said he should have mentioned that the last page of the booklet was included to get their personal reactions to some statements in order to see if there was a relation between personal reactions and memory. Some of the subjects invariably turned to the last page, and the experimenter immediately cautioned everyone not to look at the last page until they had answered all of the memory items and not to return to the memory items once they had begun to give their personal reactions. The last page contained 10 Likert-type items concerning the proposed bill to triple trucking license fees, with seven alternatives from strongly agree to strongly disagree. Five of the items favored increasing license fees and the other five opposed increasing license fees. A measure of favorability to the proposed bill was calculated by assigning scores from +3 to −3 to the alternatives for each item and then summing over the 10 items (the more positive the score, the greater the favorability to increasing license fees).

After the subjects had answered all the items in the booklet, the experimenter asked them to write a few sentences on the back describing their reactions to the study. He told them they could say anything they wanted about the study. The comments of one male subject in each condition indicated suspicion about the procedure; they were excluded from the analysis of the results. Finally, before dismissing the subjects, the experimenter entreated them not to discuss the experiment with anyone.

RESULTS

One male subject in the desirable condition failed to answer the opinion items. The exclusion of this subject, together with those previously mentioned, left a total of 51 male and 15 female subjects in the desirable condition and 57 male and 15 female subjects in the undesirable condition.

Before presenting the results for the measure of agreement with the communicator's position, it should be mentioned that there was no difference between the two experimental conditions in recall of the content of the speech. The mean number of items correct on the memory test was 16.8 for the desirable condition and 16.5 for the undesirable condition. The difference did not approach significance.[2]

From the hypothesis it was expected that agreement with the communicator would be lower in the desirable condition than in the undesirable condition. The mean for the measure of favorability to increasing trucking license fees was 3.0 for the desirable condition; it was 10.8 for the undesirable condition. The difference is in the predicted direction and highly significant ($p < .001$).

The hypothesis was based on the assumption that a communicator will be perceived as less sincere and more ingratiating when the audience feels that he thinks his communication will be desirable to the audience that is addressed than when they feel that he thinks it will be undesirable to the audience. Evidence supporting this assumption comes from the ratings of the communicator's characteristics in the desirable and undesirable conditions. The communicator was rated as significantly less sincere ($p < .001$), less honest ($p < .001$), less impartial ($p < .02$), and more cynical ($p < .01$) in the desirable condition than in the undesirable condition. He was rated as more opportunistic ($p < .001$), more obliging ($p < .001$), and more tactful ($p < .001$) in the desirable condition. The only other characteristic for which there was a significant difference between the conditions was likable; the communicator was rated as less likable in the desirable condition ($p < .02$).

DISCUSSION

The results strongly confirm the hypothesis that a communication will be less persuasive when the audience feels that the communicator thinks the communication will be desirable to the audience that will receive it than when the audience feels that the communicator thinks it will be undesirable to the audience. This hypothesis provides an explanation for the previously un-explained findings by Walster and Festinger (1962) and Brock and Becker (1965) that an overheard communication was more persuasive than a regular communication only when the topic was involving and the conclusion desirable. A person who receives a desirable communication and feels the communicator is aware that he is receiving it is also likely to feel that the communicator thinks it will be desirable to him. However, there is no reason for a person who overhears a desirable communication to feel that the communicator thinks it will be desirable to him.

---

[2] Statistical significance was determined by two-tailed t tests.

Since this experiment did not include a neutral condition, in which subjects are induced to feel that the communicator thinks the communication will be neutral to the audience, one cannot tell from the results whether the difference in opinion change was due to a decrease in acceptance of the communication in the desirable condition or an increase in acceptance in the undesirable condition or both. The first possibility is more in line with the results of the studies on overheard communications. If the perception that the communicator thinks the communication will be undesirable to the audience increases acceptance of the communication, then an overheard communication should be *less* effective than a regular communication when the conclusion is undesirable. However, in Brock and Becker's (1965) study there was no difference between an overheard communication and a regular communication when the conclusion was undesirable.

In this study the audience which the communicator was supposedly addressing was not the same as the actual audience. It seems reasonable that if it were the same, the effect on opinion change of the audience's feeling about how desirable the communicator thinks the communication will be to the audience should be even stronger in the same direction. Of course, if the audience's perception of how desirable the communicator thinks the communication will be is varied through differences in the actual desirability of the communication, the effect might be weaker, since desirable conclusions tend to be accepted more readily than undesirable conclusions.

The communicator in this study was running for election, and thus he had an obvious reason to want to please the audience he was addressing. However, the hypothesis does not require that the communicator's reason be obvious. The hypothesis assumes only that the audience think that the communicator might possibly have some reason to want to please those who will receive his communication. There are very few situations for which this assumption would not be appropriate. Rarely is it completely clear that a communicator has no possible reason to want to please his audience.

Finally, it should be noted that, while the results are quite sensible, they are not obvious. A plausible case could have been made for predicting exactly the opposite difference in opinion change. One might have argued that in the undesirable condition the speaker would be regarded as a very poor politician, and so he would be perceived as less competent and hence less credible in the undesirable condition than in the desirable condition.

REFERENCES

BROCK, T. C., & BECKER, L. A. Ineffectiveness of "overheard" counterpropaganda. *Journal of Personality and Social Psychology,* 1965, **2**, 654–660.

WALSTER, E., & FESTINGER, L. The effectiveness of "overheard" persuasive communications. *Journal of Abnormal and Social Psychology,* 1962, **65**, 395–402.

(Received May 15, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 102–106

SOME SITUATIONAL CORRELATES OF RECOGNITION-RESPONSE LEVEL

EDMOND MARKS

Georgia Institute of Technology

The effects of 2 competing incentives—1 for premature response, the other for delayed response—upon the amount of cue sampling undertaken prior to making a recognition response were examined. As predicted, the incentive for early response resulted in premature termination of cue sampling and a recognition response based on only partial cues. A significant interaction effect was obtained which was interpreted as an enhancement of the effect of the incentive for premature response upon the amount of stimulus sampling by the accurate response incentive. This result was discussed within a dissonance-theory framework.

Theoretically, the tendency to utilize incomplete and thus inadequate information in making a perceptual or cognitive response can be attributed both to certain characteristics of the individual and to properties of the situation. Although some effort has been directed toward exploring the personality dynamics underlying premature response (Binder, 1958; Messick & Hills, 1960; Zuckerman & Buss, 1960), little has been done in studying the effects of environmental and instructional variables on motivation and performance on a given perceptual task. The present study was intended to examine the effects of two classes of competing incentives upon performance on a recognition-response task.

Bruner (1957), in his extensive discussion, describes the perceptual or cognitive act as one whereby "whatever is perceived is placed in and achieves its 'meaning' from a class of percepts with which it is grouped [p. 124]." This categorization is effected by the sampling of sensory input, an object or event being a member of a preestablished class if it possesses the set of attributes characteristic of that class (Binder, 1955.) Two important properties of the Bruner model are (*a*) the treatment of the perceptual/cognitive act as one continuous process over time (Bruner, 1957; Smith, 1957), and (*b*) the introduction of expectancy, hypothesis, or belief as mediating the perceptual and cognitive response (George & Handlon, 1957). These properties are combined in the following outline (due to Bruner) of the perceptual act.

The stages of perception and cognition are (*a*) an initial or precursory stimulus input which transforms stored memories into expectancies relevant to the stimulus conditions, (*b*) cue search or scanning of the environment for additional information with which to make a tentative categorizing response, (*c*) confirmation check or narrowing of the search to cues relevant to a given expectancy, and (*d*) confirmation completion at which time incongruent cues are gated out, and the confirmed expectancy enters consciousness as a stable perceptual organization. Within this framework the continuous stimulus input serves both to elicit and/or to confirm new expectancies and disconfirm old expectancies.

In the ideal case, an individual in a perceptual recognition situation makes a series of discrimination responses based on the increasing stimulus input and responds with a categorizing response when he has sampled enough cues to provide for an errorless identification. The amount of stimulus input sampled should depend solely on the amount needed for an errorless categorizing response. Both Binder (1958) and Miller (1951), however, have hypothesized that some people exhibit a tendency to make a categorizing response before they have sampled enough stimulus information to allow for a unique or correct response. Binder (1958) has termed the level of informational input at which an individual is willing to make a categorizing response the "recognition response level." Clearly, the premature termination of this sampling process increases the probability of an incorrect recognition response. Both Binder (1958) and Messick and Hills (1960) have demonstrated the reliability and pervasiveness of this tendency to utilize inadequate information.

Some data are available from studies examining selected personality correlates of recognition-response level, for example, impulsivity and intolerance of ambiguity (Binder, 1958; Marks, 1966; Zuckerman & Buss, 1960), but these results are somewhat conflicting. They are further clouded by subtle differences in the theoretical development, samples and measures employed, and the exploratory nature of these attempts. As such, the role of personality, ability, and cognitive variables in the cue-sampling process have not been adequately explored.

But personality variables represent only one possible source of variance in recognition-response level. Another class of relevant antecedents is that relating to circumstances external to the individual. Certain conditions, such as reward for early recognition, could be expected to lead to a premature termination of stimulus sampling and a recognition response based upon only partial cues. Conversely, other conditions involving incentives for correct response could lead to a delayed or overdetermined recognition response, that is, conditions leading to prolonged or exhaustive cue sampling.

In the present study, the separate and joint effects of two competing incentives—one for early recognition, the other for delayed or correct recognition—upon the amount of stimulus information sampled before making a recognition response were investigated. It was hypothesized that the main effects of the early-response and the correct-response conditions would involve a decrease and increase in scores on the recognition-response-level task, respectively, and that the mean recognition-response-level score under the simultaneous presentation of the two incentives would not differ significantly from the expected value of scores obtained under the absence of these incentives. This latter hypothesis implied that the antagonistic response tendencies are of equal strength, and that the effects of the two incentives balance out resulting in an average amount of cue sampling.

METHOD

Subjects

The subjects were randomly selected from the freshman class entering the Pennsylvania State University in the fall term of 1965. There were 772 subjects; 600 males and 172 females. The median age was 17.5.

Procedure

Recognition-Response level. The measurement of recognition-response level consisted of two phases: (*a*) a learning phase during which the subject learned the names of eight line-drawn figures, and (*b*) a test phase during which the subject was exposed to a series of drawings containing parts of one of the eight figures for which the names were learned.

For the first phase, photographic slides were made of the eight figures developed by Binder (1958). For each figure there were two slides. The first slide of each pair which contained only the figure was exposed for 4 seconds. The second slide of the pair contained both the figure and the name identifying it. This slide was exposed for 2 seconds. The eight figures were shown once in the following manner. First each figure was presented alone and then with the name identifying it in order to acquaint the subjects with the figures. Following this initial exposure the series of figures was shown in random order for 10 trials. Now when each figure appeared the subject was to write the name of the figure in a test booklet provided before the name of that figure appeared.

In the second phase, the subject was presented with a set of six specially constructed test booklets. Each test booklet was approximately 6 inches square and consisted of five pages. Special care was taken in construction of the booklets so that the subject could not see any given page unless he first tore off the page preceding it.

On the first page was the identifying number of the booklet and also space for the subject's name. Pages 2 through 5 contained drawings which represented parts of one of the eight figures learned in the first phase. The drawings ranged from the most incomplete representation of the figure (Page 2) to the complete figure (Page 5) and were so constructed that each drawing uncovered in series reduced the uncertainty associated with the given figure by one-half. This was effected by means of elements common to all of the eight test figures; that is, the drawing on Page 2 was common to all eight figures, the drawing on Page 3 was common to four of the eight figures, the drawing on Page 4 was common to only two of the figures, while the drawing on Page 5 was the complete figure. The eight figures and test drawings are presented in Appendix A.[1]

The subjects were given unlimited time to work on the six booklets; when they felt that they could identify the figure represented, they were to write its name on that page and not tear off any more pages. Their recognition-response-level score was the total number of pages pulled on all six booklets.

Incentive conditions. To examine the effects of the incentives for both premature and delayed recognition response upon recognition-response level, a complete two-way factorial design was employed. The subjects were randomly assigned to the four cells, 193 subjects in each cell.

---

[1] Stimulus figures and test drawings have been deposited with the American Documentation Institute. Order Document No. 9212 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Subjects in the premature-response incentive condition were told that the recognition-response-level task was a measure of perceptual skill and that their score would be inversely related to the number of pages which they tore from the test booklet. A schedule of score points for the number of pages removed was given to the subjects. The second level of this factor—absence of the incentive—contained no instructions concerning score on the task.

The subjects in the delayed-response incentive condition were required to indicate the degree to which they were certain of the correctness of their answer when making a recognition response. For this purpose a 5-point scale ranging from completely certain to completely uncertain was provided below the space used for naming the figure. It was felt that these instructions offered an incentive for accurate responding. Since the subject could not be certain of the accuracy of his response until he had removed all the pages of the test booklet, this condition would require maximum cue sampling or delayed response. Again, the second level of this factor—absence of the incentive—contained no instructions concerning accuracy of recognition response.

RESULTS

Sex Differences

Before the hypotheses relating to recognition-response level were examined, the male and female samples were tested for the equality of their respective means and variances of the recognition-response-level measure. Because the observed test criteria were not significant at the .05 level, the data from the male and female samples were pooled and further analyses conducted on the entire sample.

Reliability of the Recognition-Response Measure

The split-half reliability coefficient—corrected for length by the Spearman-Brown prophecy formula—of the recognition-response-level scores was .80. Since there were only six items (series) comprising this measure, this was considered to be an acceptable reliability level.

Incentive Effects

The results of the analysis of variance in which the two types of incentives were used as independent variables—and recognition-response level as the dependent variable—are presented in Table 1. The main effect of the incentive for premature response and the interaction effect of the two incentives were significant at the .05 level. The test criterion value associated with the main effect of the incentive for accurate response failed to reach the .05 level of significance.

TABLE 1

ANALYSIS OF VARIANCE FOR THE TWO INCENTIVE CONDITIONS

Source	*df*	*SS*	*MS*	*F*
Incentive for premature response	1	159.6	159.6	14.0*
Incentive for accurate response	1	0.3	0.3	0.0
Premature × Accurate	1	57.7	57.7	5.0*
Error	768	8806.1	11.4	

* $p < .05$.

The interpretation of the main effect of the incentive for premature recognition response was unambiguous. Under such incentive conditions, the mean recognition-response levels were consistently lower than those recorded when this incentive was absent, regardless of the level of the incentive for accurate response. The role of incentive for accurate response upon recognition-response level was less obvious and could not be interpreted independently of the incentive for premature response. The relationships between incentive for accurate response and recognition-response level—plotted separately for the levels of premature-response incentive—are provided in Figure 1. As hypothesized, the incentive for accurate response led to an increased mean recognition-response level when the incentive for premature response was absent. Quite unex-

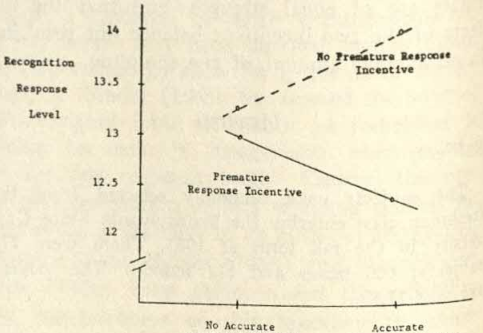


FIG. 1. The relationship of incentive for accurate response to recognition-response level under the presence (circle on solid line) and absence (circle on broken line) of incentive for premature response.

pected, however, was the noticeable drop in mean recognition-response level when both incentives were presented simultaneously, since these two conditions were hypothesized as having antagonistic but counterbalancing effects upon performance. The disordinal interaction of the two incentives, however, washed out any significant main effect relating to the incentive for accurate response.

DISCUSSION

As hypothesized, the introduction of an incentive for premature recognition response resulted in curtailment of cue sampling, despite the increased probability of such responses being inaccurate. Subjects were apparently willing to utilize inadequate and partial information in making a recognition response in an effort to improve their scores on a measure of perceived skill. The effect of this particular incentive was, however, less pronounced when the subjects were *not* provided with an incentive for accurate recognition response, the mean difference here being less than .5 page torn off. The significant "disordinal" interaction (Lubin, 1961) of the two incentive effects—aside from yielding the interpretation of the "nonsignificant" accuracy incentive main effect questionable—raised the interesting problem of one of two competing incentives facilitating rather than inhibiting the effects of the other upon performance.

It is apparent from Figure 1 that the making of an evaluative response regarding a previous recognition response, that is, incentive for accurate response, leads to increased cue sampling but only when the incentive for premature response is absent. Most interesting, however, was the joint effect of the two incentives upon recognition-response level. Contrary to expectation, the mean recognition-response level for this combination was lower (12.4) than the mean for the premature-response-incentive–no-accurate-response-incentive condition (12.9). Dissonance theory (Festinger, 1957) provides one possible explanation for these results. Subjects in the combined incentive condition were hypothesized to operate under two conflicting or competing response tendencies—one is generated by reward toward premature response; the other, by accuracy conditions toward delayed response. The apparent resolution in favor of the premature-response incentive has, itself, motivational properties (Brehm & Cohen, 1962) in that the subject once having made the premature recognition response must still handle the dissonant cognitions associated with making a certitude response. Within the dissonance framework, the tension associated with this dissonant state can be reduced by engaging in even greater premature responsivity, thus reducing the cognized intensity of the certitude motivation. Some support for this position was obtained from the systematic decrease in the mean number of pages removed over the six trials under the joint incentive conditions, a result not observed in any other joint condition.

This apparent facilitation of the effects of the premature-response incentive upon recognition-response level by the incentive for accurate response raises certain questions concerning the relationship between parameters of the incentives and the obtained results. We might ask what effects such variables as size of incentive or subjective utility of an outcome, for example, a rapid response or accurate response, have upon these results. It may well be that increasing the subjective utility of an accurate response while holding the premature-response incentive constant would wash out the presently observed interaction effect and possibly even the main effect associated with the premature-response incentive. Nonetheless, the present results indicate that under certain conditions incentives for competing response tendencies may interact to the extent that one incentive accentuates or makes the effects of another incentive upon the amount of cue sampling more pronounced.

REFERENCES

BINDER, A. A statistical model for the process of visual recognition. *Psychological Review,* 1955, **62,** 119–129.

BINDER, A. Personality variables and recognition response level. *Journal of Abnormal and Social Psychology,* 1958, **57,** 136–142.

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance.* New York: Wiley, 1962.

BRUNER, J. S. On perceptual readiness. *Psychological Review,* 1957, **64,** 340–358.

FESTINGER, L. *A theory of cognitive dissonance.* Stanford: Stanford University Press, 1957.

GEORGE, F., & HANDLON, J. A language for perceptual analysis. *Psychological Review,* 1957, **64,** 14–25.

LUBIN, A. The interpretation of significant interaction. *Educational and Psychological Measurement,* 1961, **21,** 807–817.

MARKS, E. Some personality and situational correlates of recognition response level. Unpublished manuscript, Pennsylvania State University, Student Affairs Research, Report 66-4, 1966.

MESSICK, S., & HILLS, J. R. Objective measurement of personality: Cautiousness and intolerance of ambiguity. *Educational and Psychological Measurement,* 1960, **20**, 685–699.

MILLER, J. G. Unconscious processes and perception. In R. Blake & G. Ramsey (Eds.), *Perception, an approach to personality.* New York: Ronald Press, 1951. Pp. 258–282.

SMITH, G. Visual perception: An event over time. *Psychological Review,* 1957, **64**, 306–313.

ZUCKERMAN, M., & BUSS, A. Perceptual defense and "prerecognition responsivity" in relation to hostility, anxiety, and impulsivity. *Journal of Clinical Psychology,* 1960, **16**, 45–50.

(Received May 19, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 106–108

PAIN TOLERANCE AND GROUP IDENTIFICATION[1]

ARNOLD H. BUSS AND NORMAN W. PORTNOY
Rutgers—the State University *University of Pittsburgh*

Groups were scaled for strength of identification, and high, middle, and low strengths were selected for an experiment. Men were tested for tolerance of pain induced by electric shock. Between the 1st and 2nd pain-tolerance trials they were given false norms about the pain tolerated by their reference group. These false norms increased pain tolerance: the greater the strength of identification, the greater the increase in pain tolerance. In addition to reference groups, comparison groups influenced the amount of change in pain tolerance.

Several years ago Lambert, Libman, and Poser (1960) investigated pain tolerance in relation to religious identification. When Jewish women were told that Jews have less tolerance for pain than non-Jews, the subjects tolerated significantly more pain. When Protestant women were told that Protestants have less tolerance for pain than non-Protestants, the subjects did *not* tolerate significantly more pain. However, the pain tolerance of Protestant women increased significantly when they were told that Christians tolerate less pain than Jews.

These findings may be explained in terms of group identification. To identify oneself as a Jew or Protestant implies some pride in the religious group, feeling a part of it, being involved in its reputation, taking pride in its accomplishments, defending it against attack, etc. Identification with a group implies that what affects the group must affect its members. When an individual perceives that his group's reputation is being questioned or threatened, he will defend the group even at the cost of personal discomfort. This tendency to defend or work for a group or team goal is strongly rewarded in our culture. Thus it is not surprising that women take more pain when told that their religious group is "inferior" in pain tolerance; their increased tolerance of pain is presumably a behavioral outcome of their identification with the religious group.

This explanation, which accounts for the Lambert et al. findings, applies to all group identification, not merely to religious identification. Furthermore, if group identification leads to increased tolerance of pain, it seems reasonable to assume that the relationship is quantitative: the stronger the group identification, the greater the increment in pain tolerance. The present study sought to test this hypothesis.

Another purpose of the study was to investigate the *comparison* group. Lambert et al. (1960) showed that Protestant women increased pain tolerance when Christians were compared to Jews but not when Protestants were compared to non-Protestants. Presumably, mentioning a specific, rivalrous comparison group (Jews) rendered the group identification salient, whereas leaving the comparison group vague (non-Protestants) did not. The present study pursued this question by using both strong and weak rivals as comparison groups.

METHOD

Scaling

The initial problem was to discover groups that varied in the strength with which members identify.

---

[1] This study is an adaptation of a master's thesis by the junior author under the supervision of the senior author, who is responsible for this report. We acknowledge the suggestions of Bertram Cohen and Martin Manosevitz concerning the manuscript.

Of the many groups available to college-age subjects, eight were selected for scaling. The eight groups were presented in alphabetical order to groups of subjects who were instructed to rank them for strength of identification: how strongly committed they felt toward the group, how important it was to belong to the group, or how strongly they felt as members of the group. The groups American, gender, religion, college, and state needed no elaboration. Concerning vocation, subjects were told to consider their planned career; concerning age, they were told to contrast their age group with children, the middle-aged, and the elderly; concerning club, they were told to consider sororities, fraternities, or any nonprofessional organization.

The mean rankings are presented in Table 1. Although the relative positions of the groups may be of interest in themselves, the focus here is on the selection of groups that are high, middle, and low in strength of identification. Two criteria were used in selecting groups. The first was that the men's and women's ranking be similar, so that any findings that emerged would ostensibly apply to both sexes. The second was that each reference group include a single class of individuals; for example, American refers to a single class, defined by citizenship. Such groups as religion and vocation include many classes of individuals (at least three major religions and possibly hundreds of vocations).

Design

On the basis of these criteria, three *reference groups* were selected: American (high), gender (middle), and college (low). In addition to reference groups, there were also to be *comparison groups;* the latter were selected on an a priori basis. For the American reference group the strong comparison group was Russian, and the weak comparison group was Canadian; the reasons appear obvious. For the college reference group (University of Pittsburgh) the strong comparison group was Pennsylvania State University, and the weak comparison group was Carnegie Institute of Technology. Penn State is a traditional athletic rival, whereas Carnegie Tech is not. For gender, only one alternative was available as a comparison group.

The experimental design is as follows. There are three reference groups, and they vary in strength of identification. Two of the reference groups have both strong and weak comparison groups. The third reference group was men because only men were used as subjects. Women were assigned as a strong comparison group on the assumption that there is some truth in the notion of "the war of the sexes."

Thus there were five experimental groups. The sixth group was a control. The N for each group was 15 men, yielding a total N of 90.

TABLE 1

MEAN RANKINGS OF GROUP IDENTIFICATION

Reference group	Men[a]	Women[b]
American	2.3	2.5
Gender	3.3	3.2
Religion	3.4	3.0
Vocation	4.3	3.0
Age	5.0	5.2
College	5.3	5.0
Club	5.7	6.4
State	6.5	6.2

[a] $N = 167$.
[b] $N = 186$.

Apparatus and Procedure

The pain was caused by electric shock, delivered through a finger electrode. Electrode paste was used to reduce individual differences in pain tolerance. The paste increased the current readings by a factor of 5. This caused no concern because an own-control design was employed, the dependent variable being the *change* in amperage from the first to the second trial.

The following instructions were read to each subject:

> You will be given some brief electric shocks. The purpose of this study is to standardize pain tolerance norms. The shock will be administered through this finger [electrode placed on fourth finger of left hand]. The shocks will be short and any sensations you feel will cease as soon as the current is turned off.
>
> I want you to tell me when you first feel any sensation, when the shock first begins to hurt you, and when the pain is such that you wish the shock to be discontinued.

Trial 1 then began, with the shock level starting at 0 volts and increasing in 5-volt steps. Shock duration was 1/4 second. Threshold, pain threshold, and pain tolerance were recorded in milliamperes of current, but only the pain-tolerance data are relevant here.

During a 3-minute rest period, subjects in each of the experimental groups were given the false norm appropriate to the experimental condition: on the average Russians have a greater tolerance for pain than Americans, women stand more pain than men, etc. Otherwise the conversation was general. No false norm was given in the control group. Then Trial 2 began and again the three measures were recorded.

RESULTS

The dependent variable was a change score: milliamperes on Trial 2 minus milliamperes on Trial 1. The data for the six groups of subjects are presented in Table 2. The left-hand column lists the change scores of the strong comparison groups and the control group. Note that the groups line up in order of strength of identification: the stronger the ranked identification, the greater the increase in pain tolerance. Subjects in the control group showed a slight decrease in

TABLE 2

MEAN CHANGE IN PAIN TOLERANCE[a]

Reference group	Strong comparison	Weak comparison
American	3.9	0.9
Gender	2.8	
College	1.9	1.1
Control	−0.3	

[a] In milliamperes.

pain tolerance, as did the control subjects of Lambert et al.

The significance of these trends was evaluated by analysis of variance of the data of these four groups. The F of 9.7 was significant at the .001 level ($df = 3, 56$). All three strong comparison groups were significantly different from the control group, the p values of the t tests all being less than .001. Among the experimental groups, the American-Russian and Pitt-Penn State groups differed significantly ($t = 2.2$, $p < .05$). The general trend of the findings is clear: group identification does increase pain tolerance, and the stronger the identification, the greater the increase.

The right-hand column of Table 1 lists the change scores of the weak comparison groups. The increase in pain tolerance for both of these groups was significantly greater than that of the control group (both t's were significant beyond the .01 level).

Next, let us compare the strong comparison groups with their appropriate weak comparison groups. For both the American and the college reference groups, the strong comparison group (Russian or Penn State) yielded a greater increase in pain tolerance than the weak comparison group (Canadian or Carnegie Tech). Only the Russian versus Canadian difference was significant ($t = 3.0$, $p < .01$). The Penn State versus Carnegie Tech difference was smaller and not significant, which may be attributed to the weak identification of college (the University of Pittsburgh).

DISCUSSION

The findings of this experiment clearly support the hypothesis that there is a functional relationship between strength of group identification and pain tolerance: the higher the rank of a group in strength of identification, the greater the increase in pain tolerance. These results confirm and extend the findings of the Lambert et al. (1960) study, which used religious groups and women subjects. The present research used men as subjects and nonreligious reference groups. One difference in method deserves comment. In the present study groups were first ranked for strength of identification, and these ranks were then used to vary strength of identification in the pain-tolerance experiment. The use of a scaling procedure to obtain differences in strength of identification marks an advance over the Lambert et al. study in that it permits quantifying the relationship between group identification and its behavioral manifestations.

The strength of the comparison groups also proved to be important. Pain tolerance increased more with a strong comparison group than with a weak comparison group, although the difference was significant only with the strong reference group (American). Why do comparison groups exert this influence? There appear to be two reasons. First, one determinant of strength of identification is the salience of the group, that is, its distinctiveness from other groups. The greater the difference between reference group and comparison group, the more likely the latter will be labeled as strong. Thus Russia, which has a different language, political system, and geography, is more distinct and therefore a stronger comparison group for America than is Canada, which has a similar language, political system, and geography. Presumably, the more distinct (and therefore the stronger) the comparison group, the greater is the individual's feeling of being a member of the reference group; that is, the more salient is group identification.

The second reason for the influence of comparison groups appears to be competition. The more competitive two groups are (e.g., America versus Russia), the stronger will be one as a comparison group for the other. Competition is a strong motivator and might easily lead to an increase in pain tolerance.

In the present study, strong and weak comparison groups were chosen a priori, after an informal consensus of colleagues. No one would seriously question the placement of Russia as a strong comparison group and Canada as a weak one, and certainly the experimental findings would appear to sustain the selection. However, in future research it would be preferable to have comparison groups scaled in the same way that reference groups were scaled. Such scaling would presumably lead to quantitative and more detailed findings concerning the behavioral manifestations of group identification.

REFERENCE

LAMBERT, W. E., LIBMAN, E., & POSER, E. G. The effect of increased salience of a membership group on pain tolerance. *Journal of Personality,* 1960, **28**, 350–357.

(Received May 20, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 109-114

DECODING OF INCONSISTENT COMMUNICATIONS[1]

ALBERT MEHRABIAN AND MORTON WIENER

University of California, Los Angeles *Clark University*

The present study dealt with inconsistent communication of attitude in 2 components of a message. 3 degrees of attitude (positive, neutral, and negative) communicated in single-word contents were each combined with 3 degrees of attitude communicated in tone of voice. It was found, consistent with the proposed hypothesis, that the variability of inferences about communicator attitude on the basis of information available in content and tone combined is mainly contributed by variations in tone alone. For example, when the attitude communicated in content contradicted the attitude communicated by a negative tone, the total message was judged as communicating a negative attitude. The limitations of the findings, as well as their implications for the double-blind theory of schizophrenia, were discussed.

The present study is concerned with one variant of "implicit attitude" communication. The attitudes under consideration are affective or evaluative in character. One variant of implicit attitude communication, that is, vocal (e.g., tone of voice) and nonverbal (e.g., facial expression, gesture, or postural orientation), has been the object of many studies. Reviews of the literature by Davitz (1964), Mahl and Schulze (1964), Starkweather (1964), and Marsden (1965) indicate that subjects can make reliable judgments of feelings or attitudes on the basis of vocal and nonverbal stimuli. A second variant of implicit attitude communication involves the simultaneous and inconsistent communication of attitudes in two or more components of a message, with each component independently communicating a different degree of positive-negative attitude.

There are no reported experimental investigations of multicomponent inconsistent communications of attitude in which the independent and combined effects of the components were assessed. However, there are a few studies in which consistent two-component communications were investigated. For example, Levitt (1964) and Williams and Sundene (1965) investigated the independent and combined effects of facial and vocal communication of emotions. In the latter study, it was found that all three modes of facial, vocal, and combined facial-vocal communication are recognized in terms of the three factors of general evaluation, social control, and activity.

In view of the frequent co-occurrence of verbal and nonverbal cues in most communication situations, the understanding of attitude communication requires a study of the interactive effects among such cues. The present experiment was intended as a step in this direction. In the experiment, quantification of the degree of inconsistency between components in a communication was possible only when the two components could be scaled along a common dimension. Thus, the general evaluation dimension obtained in the Williams and Sundene (1965) study was used since it seemed appropriate for the study of attitude communication. The two components were tone and content in a verbal message. Tone refers to those permissible variations of pitch and stress which are linguistically unnecessary; content refers to word meanings.

While the independent effects of content can be assessed through the use of written communications, the effects of tone are somewhat more difficult to obtain. Davitz (1964) reviewed the experimental literature relating to the communication of feelings through tonal variations. In the studies reviewed, the independent effects of tone were obtained either through the use of electronic filters which rendered the content portions of a message incomprehensible, or through the use of the alphabet or numbers as carriers of tone. The former method has the disadvantage of modifying the tonal component of a message in the process of eliminating the content component, while the latter method has the disadvantage of restricting the range of contents on which tone may be superimposed. The methodology of the present experiment was designed to try to circumvent these problems so

---

[1] This study was supported in part by University of California Grant 2189 and in part by United States Public Health Service Grant M-3860. The authors wish to thank Sylvia Norman and Dena Sugar, who acted as speakers in the recording of the experimental conditions, Allen Lowe, who assisted in running the subjects, and Susan Ferris, for her comments on the manuscript.

that the same verbal materials could be used to assess the independent effects of tone, the independent effects of content, and the combined effects of tone and content.

Three degrees of attitude (positive, neutral, and negative) in the content of the communications were each combined with three degrees of attitude in the tone of the communications. Thus, with the exception of the positive-content–positive-tone, neutral-content–neutral-tone, and negative-content–negative-tone conditions, all other six conditions involved inconsistency between the tonal and content components of the message. Extrapolating from informal observation of responses to sarcastic attitude communications (i.e., the dominance of the tonal component over the content component), it was expected that in these six instances the effects of tone rather than those of content would determine the addressee's interpretation of the total message. It was therefore hypothesized that for the nine Tone × Content conditions the effects of tone are significantly greater than the effects of content, indicating that tone and content do not contribute to the interpretation of the total message in proportion to their independent effects.

METHOD

Subjects

Forty-five subjects were used in the initial selection of the positive-, neutral-, and negative-affect communicating words (i.e., contents). A second group of 30 subjects was used in the remaining portion of the study. All 75 subjects were University of California undergraduates.

Materials

The positive-, neutral-, and negative-affect communicating words were selected by the following procedure. The words DEAR, GREAT, HONEY, LOVE, and THANKS were initially selected as likely positive-attitude communicating contents. The words MAYBE, OH, REALLY, SO, and WHAT were selected as likely neutral-attitude communicating contents. The words BRUTE, DON'T, NO, SCRAM, and TERRIBLE were selected as likely negative-attitude communicating contents. A group of 45 subjects received one page of instructions and a 15-page booklet, each page of which contained one of the words noted above and a 12-centimeter scale designated "positive" and "negative" at its poles. The order in which the 15 words were presented to each subject was randomized. Also, for each word, the positioning of the positive and negative poles was counterbalanced. The instructions were as follows:

> Each of the words in your booklet can be imagined as having been said by person X to person Y (e.g., X says to Y, "later"). For each word, indicate on the scale what you think X's attitude is towards Y. On the scale, "positive" refers to liking, high evaluation, or preference for Y, while "negative" refers to disliking, low evaluation, or lack of preference for Y.

The means and sample standard deviations of judgments of degree of positive attitude were as follows on a −6 (i.e., most negative) to +6 (i.e., most positive) scale for the assumed positive words: HONEY (3.21, 2.24), THANKS (3.27, 2.02), DEAR (3.27, 2.51), GREAT (3.62, 1.90), LOVE (4.48, 1.76); for the assumed neutral content words: MAYBE (−.43, 1.94), REALLY (.39, 1.83), OH (−.04, 2.10), SO (−1.59, 1.75), WHAT (−.49, 1.37); for the assumed negative words: DON'T (−2.86, 2.24), BRUTE (−3.12, 2.12), TERRIBLE (−3.21, 2.23), NO (−2.38, 2.08), SCRAM (−4.33, 1.75). On the basis of these findings, the words HONEY, THANKS, and DEAR were selected as comparable instances of positive contents; the words MAYBE, REALLY, and OH were selected as comparable instances of neutral contents; and the words DON'T, BRUTE, and TERRIBLE were selected as comparable instances of negative contents.

Two female speakers were employed to read each of the nine selected words with each of the three positive, neutral, and negative tones. For the positive-, neutral-, and negative-tone conditions, the speakers were instructed to say the words, irrespective of contents, in such a way as to convey an attitude of liking, high evaluation, or preference; a neutral attitude, that is, neither liking nor disliking; and an attitude of disliking, low evaluation, or lack of preference, respectively, towards the target person.[2] All possible combinations of two speaker conditions, three tone conditions, three content conditions, and three instances of each content condition were recorded on tape.

Procedure

A random order of the 54 conditions was administered to each subject. An individual subject received only one of the following three instructions:

> We are interested in finding out how well you can judge the feelings of a speaker towards the person whom she is addressing. You are to imagine each one of the following words as being said by one person, the speaker, to another and you are to judge what the speaker's attitude is towards her addressee [attend to content and tone instructions] using all the information available; [attend to content only instructions] using only the information contained in the meanings of the words and ignore the information contained in the tone; [attend to tone only instructions] using only the information contained in the tone and ignore the information contained in the meanings of the

---

[2] It may be indicated that the object of this experiment was not the investigation of the problem as to what types of pitch and stress patterns correspond to positive-, neutral-, and negative-attitude communicating tone. The problem was, rather, given that tone is judged as having any one of these positive-, neutral-, or negative-attitude communicating qualities, what are its interactive effects with positive-, neutral-, and negative-attitude communicating contents.

> words. Record your judgments on the answer sheet which has been provided for you. Each time you hear the speaker say a word, we would like you to judge the degree of the speaker's positive versus negative attitude towards the addressee. "Positive" refers to liking, preferring or having high evaluations and "negative" refers to disliking, not preferring or having low evaluations of the addressee.
>
> Use the scale provided on this instruction sheet to indicate your estimate of speaker attitude in each case.

A scale numbered −3 through +3, extending from a pole designated negative to a pole designated positive, was inserted at this point, and the remaining instructions explained the use of the scale and recording of judgments on an answer sheet which was also provided.

Ten subjects received the instructions to use "all the information available," 10 subjects received the instructions to use "only the information contained in the meanings of the words and ignore the information contained in the tone," and 10 subjects received the instructions to use "only the information contained in the tone and ignore the information contained in the meanings of the words." These three sets of instructions and their accompanying methodology were designed to make possible an independent evaluation of the contribution of tone and content together, content alone, and tone alone, respectively, in the same spoken medium of communication and through the use of identical materials.

RESULTS

Each subject recorded 54 responses which were three replications of each of the 2 Speaker × 3 Content × 3 Tone conditions. Thus, for each Speaker × Tone × Content combination, three measures were obtained from each subject corresponding to the three words in the particular content category involved (e.g., the words MAYBE, REALLY, and OH in the neutral content category). These response scores, which had a possible range of from −3 to +3, were analyzed in three separate analyses, one for each of the three types of instructions.

The means of the nine Tone × Content conditions obtained for Speaker A and Speaker B communications under each of three instruction conditions are given in Tables 1 and 2, respectively.

The data obtained from the subjects who were instructed to use only the information contained in the meanings of the words were analyzed in a 2 × 3 × 3 factorial design, replicated over subjects, in which there were 2 speaker, 3 tone, and 3 content conditions. There was a significant effect due to content, $F\ (2, 18) = 39.96$, $p < .01$. The effects due to tone, $F\ (2, 18) = 3.51$, $p < .1$, and Speaker × Tone, $F\ (2, 18) = 3.07$, $p < .1$, approached significance.

TABLE 1

DEGREE OF INFERRED POSITIVE ATTITUDE FOR SPEAKER A AS A FUNCTION OF THE NINE CONTENT × TONE STIMULUS CONDITIONS AND INSTRUCTIONS

Instructions	Content	Tone		
		Negative	Neutral	Positive
Use content only	Negative	−1.33	−1.00	−0.67
	Neutral	−0.47	−0.17	0.35
	Positive	1.03	1.30	1.70
Use tone only	Negative	−2.47	−0.03	1.40
	Neutral	−2.07	−0.67	1.73
	Positive	−1.37	0.17	1.63
Use tone and content	Negative	−1.77	0.30	1.21
	Neutral	−1.67	−0.40	1.10
	Positive	−0.87	0.40	1.10

For Speaker A communications, the Newman-Keuls method yielded significant differences at the .05 level for all three comparisons among means of the three content categories for each tone level. For example, the negative-tone–negative-content cell mean was less than the negative-tone–positive-content and the negative-tone–neutral-content cell means. The negative-tone–neutral-content cell mean was also less than the negative-tone–positive-content cell mean. Similarly, judgments based on Speaker B communications yielded significant differences at the .05 level for all three comparisons among the means of the three content categories for each tone level. Thus, the independent effects due to content were unambiguous for both Speaker A and Speaker B communications.

A similar analysis of the data obtained from

TABLE 2

DEGREE OF INFERRED POSITIVE ATTITUDE FOR SPEAKER B AS A FUNCTION OF THE NINE CONTENT × TONE STIMULUS CONDITIONS AND INSTRUCTIONS

Instructions	Content	Tone		
		Negative	Neutral	Positive
Use content only	Negative	−1.03	−1.00	−0.80
	Neutral	0.07	0.13	0.53
	Positive	0.67	1.63	1.63
Use tone only	Negative	−1.60	0.47	0.83
	Neutral	−0.43	0.20	1.53
	Positive	−1.13	0.73	1.77
Use tone and content	Negative	−0.83	0.20	0.83
	Neutral	−0.40	0.03	0.83
	Positive	−1.03	0.87	1.43

the subjects who were instructed to use only the information contained in the tone of the words indicated significant effects due to speaker, $F\ (1, 9) = 21.18, p < .01$; tone, $F\ (2, 18) = 178.42, p < .01$; content, $F\ (2, 18) = 5.58, p < .05$; Speaker × Tone, $F\ (2, 18) = 14.54, p < .01$; Tone × Content, $F\ (4, 36) = 7.09, p < .01$; and Speaker × Tone × Content, $F\ (4, 36) = 4.50, p < .01$. Separate analyses for each speaker level indicated the following effects: For Speaker A, there was significance due to tone, $F\ (2, 18) = 191.30, p < .01$; content, $F\ (2, 18) = 6.94, p < .01$; and Tone × Content, $F\ (4, 36) = 4.61, p < .01$. For Speaker B, there was significance due to tone, $F\ (2, 18) = 77.23, p < .01$; content, $F\ (2, 18) = 3.70, p < .05$; and Tone × Content, $F\ (4, 36) = 6.48, p < .01$. Furthermore, for Speaker A communications, the Newman-Keuls method yielded significant differences at the .05 level for all three comparisons among the means of the three tone categories for each, content level. For Speaker B communications, the Newman-Keuls method yielded significant differences at the .05 level for all three comparisons among the means of the three tone categories for the neutral- and positive-content levels. For the negative-content level, however, while the negative-content–negative-tone cell mean was significantly less than the negative-content–positive-tone and negative-content–neutral-tone cell means, the latter was not significantly less than the negative-content–positive-tone cell mean. Thus, for Speaker A communications, the independent effects of tone were unambiguous. However, for Speaker B communications, the effects of tone were weaker than those of Speaker A (as indicated by the significant Speaker × Tone interaction obtained in the overall analysis), and the effects for Speaker B were significantly different in only eight of the nine comparisons of cell means.

The effects due to content which were obtained under "use tone only" instructions, although incidental, require some consideration. These effects were different for different tone levels of each speaker. Moreover, comparisons of cell means indicated only a few significant differences. For Speaker A, the Newman-Keuls method yielded three out of nine significant differences at the .05 level. One of these effects was in a direction opposite to what might have been expected: the neutral-tone–negative-content cell mean was greater than the neutral-tone–neutral-content cell mean. For Speaker B, the Newman-Keuls method also yielded three out of nine significant differences at the .05 level. Once again, one of these effects was in a direction opposite to what might have been expected: the negative-tone–positive-content cell mean was less than the negative-tone–neutral-content cell mean. In brief, the effects due to content which were obtained under "use tone only" instructions were not only weak, but also somewhat inconsistent.

Analysis of the data obtained from the subjects who were instructed to use all the information available indicated significant effects due to speaker, $F(1, 9) = 10.55, p < .01$; tone, $F\ (2, 18) = 30.57, p < .01$; Speaker × Tone, $F\ (2, 18) = 6.00, p < .01$; and Speaker × Tone × Content, $F\ (4, 36) = 4.78, p < .01$. Separate analyses for each speaker level indicated the following effects. For Speaker A, there was significance due to tone, $F\ (2,18) = 36.98, p < .01$. For Speaker B, there was significance due to tone, $F\ (2, 18) = 17.77, p < .01$; and Tone × Content, $F\ (4, 36) = 4.22, p < .01$. Furthermore, for Speaker A communications, the Newman-Keuls method yielded significant differences at the .05 level among the means of the three tone categories for each content level. For Speaker B communications, the Newman-Keuls method yielded significant differences at the .05 level among the means of the three tone categories for the negative- and positive-content levels. For the neutral-content level, however, the neutral-content–negative-tone cell mean was less than the neutral-content–positive-tone cell mean, but did not differ significantly from the neutral-content–neutral-tone cell mean, while the latter was less than the neutral-content–positive-tone cell mean.

The above results may be summarized as follows: For Speaker A communications, under "use content only" instructions, the effects of content were all significant; under "use tone only" instructions, the effects of tone were also all significant; and finally under "use all the information" instructions, the effects of tone were all significant, while there were no effects due to content. For Speaker B communications, under "use content only" instructions, the effects of content were all significant; under "use tone only" instructions, the effects of tone were somewhat weak and only eight of the nine comparisons of adjacent cell means yielded significant differences; and finally under "use all the information" instructions, the effects of tone were again somewhat weak and only eight of the nine comparisons of cell means yielded significant differences, while there were no effects due to content. Finally, under "use all the information" instructions, analysis of responses to Speaker A communications did not yield a significant Tone × Content effect, while analysis of responses to Speaker B communications did yield a significant Tone × Content effect.

DISCUSSION

A discussion of the combined effects of tone and content on inferred attitude is meaningful when the independent effects of tone and content are ascertained. As already noted, the methodology of the above experiment was designed to allow the identical stimulus materials for obtaining the independent effects of content, the independent effects of tone, and the combined effects of tone and content on inferred attitudes. The findings regarding the independent effects of content were unambiguous. When subjects were asked to attend to the content portion of a spoken verbal communication, which also included systematic variations in its tonal component, they responded primarily to the content portion. The discriminations between varying degrees of positive-negative attitude communication in content indicate that the independent effects of content were clearly operative. Furthermore, the judgments of attitude obtained under "use content only" instructions corroborated the judgments of attitude which were obtained when other subjects were presented with only the written contents.

The findings regarding the independent effects of tone were somewhat complicated by the fact that when subjects were asked to attend to the tonal portion of a spoken verbal communication, which also included systematic variations in its content component, they were somewhat influenced by variations in content. However, as has already been noted, these effects due to content were weak and inconsistent. Thus, under "use tone only" instructions, subjects' reponses indicated a stronger and more consistent set of effects due to tone than due to content. Furthermore, the discriminations of varying degrees of positive-negative attitude communication in tone indicated that the independent effects of tone were clearly operative for Speaker A communications and for most of Speaker B communications (eight out of nine assumed differences).

The findings also indicated that the independent effects of tone, overall, were stronger than the independent effects of content. Thus, if tone and content were to have contributed to inferred attitudes in proportion to their independent effects, analysis of the data obtained from the subjects who were instructed to use all the information available would have yielded main effects due to both tone and content, with a stronger main effect due to tone. The actual findings did not support an assumption of proportionate contribution of tone and content to inferred attitude. The responses of subjects who were instructed to use all the information available indicated no main effects of content. The variability of response to tone and content combined was primarily determined by variations in tone. In the case of Speaker B communications only, there was, in addition to the main tone effect, a Tone × Content interaction effect. The latter effect seems to be due to the pronounced effect of negative tone in the positive-content condition.

While the above findings provided support for the hypothesis, the methodology which was used for assessing the independent effects of tone failed to solve the problem for which it was intended. It will be recalled that the methodology was designed to make possible the use of identical spoken materials for assessing the independent effects of tone, the independent effects of content, and the combined effects of tone and content. An alternate methodology could have employed written communication for assessing the independent effects of content and electronically filtered speech (with the content rendered incomprehensible) for assessing the independent effects of tone. The latter method, as well as the one used, introduces some ambiguity in the assessment of the independent effects of tone. A comparison is needed for the independent effects due to tone of a given set of spoken communications obtained with "use tone only" instructions and with the filtering technique. If the two methods yield equivalent results, they can be used in combination in an attempt to replicate the findings of the above experiment.

Having commented on some methodological problems, it is possible to proceed to a discuscion of the implications of the findings. The results indicate that judgments of attitude from "inconsistent" messages involving single words spoken with intonation are primarily based on the attitude carried in the tonal component. In other words, the tonal component makes a disproportionately greater contribution to the interpretation of the total message than does the content component. These findings regarding the relative contribution of the tonal component of a verbal message can be safely extended only to communication situations in which no additional information about the communicator-addressee relationship is available (as was the case in the above experiment). For example, it seems that communications accompanied by the communicator's commitment to action will be judged according to that action. "I hate you," said in a positive tone of voice, may be interpreted as a positive-attitude communication in the absence of additional information. However, if the communicator says, "I hate you," in a positive tone while striking his addressee, the attitude communicated is a negative one. In such instances, the positive tone indicates the communicator's pleasure at what he is doing. Information about commitments of a communicator can also be given in the verbal portion of a communication. For example, a boss can tell his assistant that he is fired in a pleasant tone of voice. Here again, the tone of voice indicates the boss' pleasure about his communication, rather than a positive attitude towards the assistant.

While the above experiment dealt with the relationship between tone and content in attitude communication, it is possible that similar relationships exist between other implicit communications of attitude (e.g., facial expression) and content. In other words, when an implicit communication of attitude is inconsistent with an explicit content communication of attitude, the contribution of the implicit component may be disproportionately greater than its independent effect. For example, smiling at the addressee while insulting him verbally generally communicates an affectionate rather than an angry attitude, in the absence of other communicator commitments to action consistent with the content portion.

Finally, the above study has some implications for the "double-bind" theory of schizophrenia. It has been hypothesized that high incidence of inconsistent communications in the families of schizophrenics compared to families of normals contributes to the generation and maintenance of the maladaptive patterns of interpersonal functioning observed in schizophrenics (e.g., Bateson, Jackson, Haley, & Weakland, 1956; Haley, 1963; Weakland, 1960). The assumptions involved in such theorizing seem to be that: (*a*) Inconsistent components in a message are decoded separately; (*b*) a consensual or adaptive response to all components would lead to incompatible responses—double bind; (*c*) the addressee or target of *frequent* inconsistent double-bind messages learns to respond with his own inconsistent messages—namely, his maladaptive responses. Assumptions *a* and *b* imply that the inconsistency between two components in a communication remains unresolved for the addressee and that there is little subordination (or superordination) of the inconsistent components with respect to each other. The present study of decoding of inconsistent communications by normals does not support these two assumptions. It is of interest to determine whether persons classified as schizophrenics, when faced with inconsistent communications, would judge the overall communications in the same ways as the subjects in the above experimental sample. If so, the higher incidence of inconsistent communication in families of schizophrenics than in families of normals could reflect a higher incidence of negative-attitude communications, provided it can be assumed that the implicit components of the reported inconsistent communications typically communicate a negative attitude. The preceding assumption is supported by the anecdotal data reported by investigators of double-bind-related phenomena (e.g., Weakland, 1960). Given the proposed interpretation of inconsistent communications, it could be argued that unusually frequent negative-attitude communicating messages do contribute to severe psychopathological functioning, but for reasons other than those given in double-bind-type conceptualization. For example, indiscriminate negative reinforcement is not conducive to learning the numerous interpersonal and social skills which are lacking in individuals classed as schizophrenics. In short, it is suggested that the relative status of various components of a message needs to be explored so that response to the total message can be specified on the basis of response to its components. With such information available, the contribution of an unusually high frequency of inconsistent messages to psychopathological functioning may be elaborated in terms of generally established theoretical concepts.

REFERENCES

Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. Toward a theory of schizophrenia. *Behavioral Science,* 1956, 1, 251–264.

Davitz, J. R. *The communication of emotional meaning.* New York: McGraw-Hill, 1964.

Haley, J. *Strategies of psychotherapy.* New York: Grune & Stratton, 1963.

Levitt, E. A. The relationship between abilities to express emotional meanings vocally and facially. In J. R. Davitz (Ed.), *The communication of emotional meaning.* New York: McGraw-Hill, 1964. Pp. 87–100.

Mahl, G. F., & Schulze, G. Psychological research in the extra-linguistic area. In T. A. Sebeok, A. S. Hayes, & M. C. Bateson (Eds.), *Approaches to semiotics.* The Hague: Mouton, 1964. Pp. 51–125.

Marsden, G. Content-analysis studies of therapeutic interviews: 1954 to 1964. *Psychological Bulletin,* 1965, 63, 298–321.

Starkweather, J. A. Variations in vocal behavior. In D. McK. Rioch & E. A. Weinstein (Eds.), *Disorders of communication.* Baltimore, Md.: Williams & Wilkins, 1964. Pp. 424–449.

Weakland, J. H. The "double-bind" hypothesis of schizophrenia and three-party interaction In D. D. Jackson (Ed), *The etiology of schizophrenia.* New York: Basic Books, 1960. Pp. 373–388.

Williams, F., & Sundene, B. Dimensions of recognition: Visual vs. vocal expression of emotion. *Audio-visual Communication Review,* 1965, 13(1), 44–52.

(Received May 24, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 1, 115–118

DOUBLE AGREEMENT AS A FUNCTION OF ITEM AMBIGUITY AND SUSCEPTIBILITY TO DEMAND IMPLICATIONS OF THE PSYCHOLOGICAL SITUATION[1]

LOREN McBRIDE AND GARY MORAN
University of Montana *University of Indiana, Gary*

The present investigation was predicated upon the assumption that double agreement (the tendency to agree to both original and reversed authoritarianism items) is a function of both S and item characteristics. Double agreement on F, Anti-Semitism, and Dogmatism scales correlated .88, .65, and .83, respectively, with the rated ambiguity of the items. This finding supports the position taken by Peabody as opposed to alternatives suggested by Rokeach and Samelson. Further, scores of Ss on the Marlowe-Crowne Social Desirability scale correlated .36, .25, and .44 with their frequencies of double agreement on the 3 measures. Both hypotheses were supported for each of the 3 tests at the .001 level of significance. Findings were discussed in terms of the role of ambiguous measures and social pressures indigenous to the testing situation as factors in the erroneous identification of authoritarianism.

One aspect in particular of the various measures employed in defining the so-called authoritarian syndrome has drawn consistent criticism. The F and Anti-Semitism scales (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), as well as the Dogmatism scale employed in a later reformulation of the authoritarian problem (Rokeach, 1956), consist exclusively of items keyed in a single direction; that is, all questions are worded such that endorsement of a statement represents a positive increment to the subject's score on the dimension being measured. Unidirectional keying of statements necessarily allows for confounding of content and response tendencies as determinants of endorsement. Moreover, confounding of content responses by response-set tendencies is particularly likely when all items are keyed positively (Berg & Rapaport, 1954). A very real question arises as to whether any authoritarianism unconfounded by acquiescence response tendencies exists whatsoever (Chapman & Bock, 1958; Christie, Havel, & Seidenberg, 1958).

Peabody (1961) presented strong evidence to the effect that while nonauthoritarians exist, authoritarians as defined by the F, Anti-Semitism, and Dogmatism scales are artifacts of acquiescence response tendencies. Original F, Anti-Semitism, and Dogmatism items were presented to his sample, and 2 weeks later these same subjects answered reversals of the original items. Peabody reasoned that the magnitiude of true authoritarianism would be reflected by the proportion of items whose original form was endorsed and reversed form rejected. For F, Anti-Semitism, and Dogmatism scales, respectively, the proportions of items meeting this criterion were .14, .10, and .15. These proportions may be compared to the magnitude of true nonauthoritarianism, that is, disagreement with original and endorsement of reversed items, as well as to the magnitude of agreement to both original and reversed items (so-called "double agreement"). For F, Anti-Semitism, and Dogmatism scales, 44%, 62%, and 45% of the responses met the nonauthoritarian criterion; 33%, 19%, and 32% of the responses are of the double-agreement type. Double agreement is about twice as frequent as true authoritarian responses. Peabody also reported that of all F, Anti-Semitism, and Dogmatism scale items agreed to in the original 70%, 63%, and 67%, respectively, were also agreed to in their reversed forms. It seems reasonable to conclude that a substantial number of authoritarians are better characterized as double agreers than as proponents of the complex ideology circumscribed by agreement with original F, Anti-Semitism, and Dogmatism scale item content.

The argument against interpreting agreement to F, Anti-Semitism, and Dogmatism scale items as evidence for the existence of an authoritarianism syndrome is based upon the incidence of double agreement. Identification of the determinants of double agreement assumes, therefore, crucial significance. In his discussions of the factors determining double agreement, Peabody (1961, 1966) proposed that acquiescence response tendencies come to the fore when an item is ambiguously worded. In his view, most authoritarianism items are ambiguous. This would explain the high incidence of double agreement in the data.

---

[1] This research is based upon a thesis directed by the junior author and submitted by the senior author to the Department of Psychology, University of Montana, in partial fulfillment of the requirements for the master of arts degree. A fuller presentation of the investigation may be found in McBride (1966).

In attempting to preserve a content view of endorsement of authoritarianism items, several alternative interpretations of the double-agreement phenomenon have been proposed. Samelson (1964) suggested that double agreement could have resulted from asymmetrical reversals of original items. If this was the case, both the original and reversed statements might lie within the agreement segment of the subject's latent attribute continuum. Double agreement in this case would not invalidate a content interpretation of agreement to original items. While logically plausible, evidence contrary to Samelson's hypothesis (Miklich, 1966; Peabody, 1966) argues against this explanation.

Rokeach (1963) has proposed yet another content interpretation of double agreement. According to this formulation, the subject "tells the truth in one case because he sees no reason why he should not; he deliberately lies in the second case because he sees a good reason why he should not tell the truth [p. 305]." Original F, Anti-Semitism, and Dogmatism scale statements are viewed by the authoritarian subject as innocuous, and, since they represent his attitudes, he endorses them. The reversals are also accepted, but not because they represent the subject's bona fide view. Rather, they are accepted because they are socially more desirable. In support of this position, Rokeach cited several anecdotes. In a more empirical vein Stanley and Martin (1964) reported a significant *negative* correlation between original Dogmatism scale items and scores on the Martin Social Desirability Scale. It would thus appear that endorsement of Dogmatism scale items is not socially desirable behavior. Further, scores on a *reversed* Dogmatism scale failed to correlate with the Martin test for a group of *high-acquiescent* subjects. The assumption that Dogmatism scale reversals are viewed as socially more desirable than original items seems untenable. Rokeach's hypothesis appears unsupported.

Rokeach (1963) also objected to Peabody's interpretation of double agreement in terms of item ambiguity. He noted that Peabody's claim was made "without providing the slightest independent empirical support, that response set is a function of ambiguity of items [p. 307]." Indeed, Peabody's (1966, p. 13) rebuttal to Rokeach's paper cited but two sources of support for the item-ambiguity hypothesis: first, that it is ". . . generally recognized by those who have examined them closely," that the items in question are ambiguous; second, that items in the F, Anti-Semitism, and Dogmatism scales have more words than those in the MMPI. That is, Peabody maintained that the MMPI is not ambiguous, that authoritarianism tests are, and that longer statements in the authoritarianism tests prove the point. Perhaps Peabody is right, but the present authors are inclined to agree with Rokeach that Peabody's evidence has not entirely settled the issue of the validity of his hypothesis. Independent, empirical evidence of the covariation of item ambiguity and double agreement on authoritarianism scales is, judging from Peabody's (1966) review, not available in the literature. The present study is predicated upon the assumption that double agreement is a function of both subject variables and characteristics of the items involved. Specifically, two hypotheses will be tested. First, double agreement will vary directly with the rated ambiguity of the paired original and reversed items. Second, double agreement will be positively correlated with the susceptibility of the subject to demand implications of the experimental situation. Implicit in this formulation is the assumption that subjects highly dependent upon social approval are differentially inclined to agree with all statements presented to them by the experimenter because they view such behavior as socially desirable.

METHOD

Materials [2]

Subjects were presented both original and reversed items used in defining the content of authoritarianism. As a detailed discussion of the nature of these stimuli is available (Peabody, 1961, 1966), only a brief description will be given here. Twenty-eight F, 20 Anti-Semitism, and 40 Dogmatism scale items were employed. Reversals of each of these items were also administered. The criterion for reversing original items was in terms of a reversal of the authoritarian aspects of the original statements rather than a mechanical reversal of item content.

---

[2] Original and reversed F, Anti-Semitism, and Dogmatism items, adopted from Peabody (1961), have previously been deposited with the American Documentation Institute. Order Document No. 6826 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.75 for microfilm or $2.50 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

Original and reversed items were presented in separate booklets. Within booklets of original items, all of the items belonging to a distinct test, for example, the F test, appeared consecutively, but were not otherwise separately identified or segregated from other items. The same was true of booklets of reversed items excepting only that within each test reversals appeared in the opposite order to the appearance of their corresponding originals. The order of appearance of separate tests was, however, counterbalanced for the sample of booklets. As far as the subject was concerned, he simply received a booklet containing 88 statements.

In addition to authoritarianism items, the Marlowe-Crowne Social Desirability (M-C *SD*) scale was administered to subjects in the testing phase of the experiment. This scale was included as a measure of differential susceptibility to social influence in general and the demand implications of experimental situations in particular (Crowne & Marlowe, 1964).

Ambiguity Ratings[3]

A separate sample of 166 psychology students, equally divided as to sex, was split randomly into two groups, one to rate the ambiguity of original, the other, of reverse items. Ambiguity ratings were made on each item separately by employing 8-point rating scales. The operational definition of "ambiguity" involves a dimension which may be summarized as the clarity of the statement. Specific written instructions to the subjects were as follows:

> You are to assess the ambiguity of the following statements in terms of the degree to which you understand their content and meaning. *We are not concerned with whether you do or do not agree with the item, nor are we concerned with whether you would admit that it is true of you.* We are solely concerned with whether or not you are clear as to what is being asked. Please rate each item's ambiguity by placing a mark at the appropriate point on the scale for each item. Two poles are presented on each scale: "clear, straightforward," and "ambiguous." You are to judge where on this continuum each item falls. Be sure and mark each statement; do not leave any out.

In this manner, mean ambiguity-scale values were obtained for each item in both original and reversed authoritarianism scales. Mean ambiguity-scale values for item pairs were calculated from these data. These item-pair means represent the ambiguity values of concern in this study.

Testing Phase

In this section of the experiment, subjects were presented either original or reversed items in their respective booklets labeled "Public Opinion Survey." The instructions at the head of the booklet stressed that interest was in the subject's personal opinion and that whether he agreed or disagreed with a particular statement there were sure to be many others of similar persuasion.[4] The subject was to indicate either agreement or disagreement with an item by appropriately marking an accompanying IBM sheet.

TABLE 1

PEARSON CORRELATIONS BETWEEN FREQUENCY OF DOUBLE AGREEMENT, ITEM AMBIGUITY, AND MARLOWE-CROWNE SCORES

Test	Item ambiguity—double agreement	M-C *SD*—double agreement
F scale	.88	.36
Anti-Semitism scale	.65	.25
Dogmatism scale	.83	.44

Note.—All probability values are associated with 1-tailed tests, $p < .001$.

The sample of 108 male and 72 female freshman psychology students was equally and randomly separated into two subconditions. One condition answered the M-C *SD* scale and then the booklets of original authoritarianism items. Two days later they answered reversals of the authoritarianism items. The other condition answered the M-C *SD* scales and reversals of authoritarianism items at the initial session, answering original authoritarianism items 2 days later. Both subconditions were, of course, combined in analyzing the data.

RESULTS

The interdependence of double agreement and the rated ambiguity of authoritarianism item pairs as well as the subjects' susceptibilities to social influence are documented in Table 1. With regard to the ambiguity hypothesis, separate correlations were computed between the frequency of double agreement to an item pair and the rated ambiguity of the item pair for F, Anti-Semitism, and Dogmatism scale items. These coefficients are .88, .65, and .83, respectively, all highly significant ($p < .001$). Clearly, the more ambiguous a pair of original and reversed items, the more likely the subject is to agree with both statements.

To test the hypothesis that subjects differentially susceptible to the demand implications of an experimental situation will tend to agree to statements presented to them by an experimenter independent of the content of these statements or the obvious inconsistencies in their responses, subjects' M-C *SD* scale scores were separately correlated with their frequencies of double agreement on the F, Anti-Semitism, and Dogmatism scales. A similar correlation was run with double agreement on all 88 items independent of the subtests to which the items belonged. Respective coefficients for the F, Anti-Semitism, and Dogmatism scales were .36, .25, and .44. For all 88 items collectively, a correlation of .44 was observed. Although the absolute magnitude of these coefficients is substantially less than was found between ambiguity and double agreement, they are nonetheless highly significant ($p < .001$).

---

[3] Gratitude is expressed to Alan J. Klockars for his valuable suggestions at this point as well as elsewhere in the execution of this experiment.

[4] Instructions adopted from Peabody (1961).

DISCUSSION

Peabody (1966) has argued that the high incidence of double agreement in his data could be attributed primarily to the ambiguous nature of items employed to measure authoritarianism. Unfortunately, he neglected to present any substantial evidence to support the claim that these items are in fact ambiguous. How ambiguous are authoriarianism items? The mean ambiguity-scale values for F, Anti-Semitism, and Dogmatism scale items in the present sample are 3.84, 4.31, and 3.77. None of these values indicate ambiguity in an absolute sense, as they all fall at or below the midpoint of an 8-point scale. Furthermore, Rokeach (1963, p. 307) found the high incidence of double agreement on the Anti-Semitism scales in Peabody's data ". . . a most curious finding. . . ." No one, in his opinion, "has even seriously claimed that A-S scale items are ambiguous," as, for example, has been claimed for F and Dogmatism scale items. The present evidence, however, indicates that the Anti-Semitism scale is significantly *more* ambiguous than either the F ($p < .02$) or the Dogmatism ($p < .01$) scales. Surely the minimal conclusion to be drawn is that a moratorium is in order on appeals to authority in defining the ambiguity of authoritarianism items.

Another point seemingly worthy of note concerns the magnitude of double agreement observed in the present sample. For F, Anti-Semitism, and Dogmatism scales, respectively, double agreement represented 24%, 12%, and 25% of the total responses. These proportions, although substantial considering the phenomenon in question, are approximately one-third lower than in the other pertinent investigation (Peabody, 1961).

With regard to issues more directly germane to the purpose of this investigation, the data clearly indicate that double agreement on authoritarianism scales is highly dependent upon the relative ambiguity of the items involved. In a practical sense, the magnitude of the correlations involved indicates that this conclusion holds with equal force for F, Anti-Semitism, and Dogmatism scales. The findings indicate a slight refinement of Peabody's hypothesis. The most ambiguous *test* does not result in the highest proportion of double agreement. Rather, the most ambiguous item pairs in each of the tests lead to the highest incidences of double agreement within the respective scales.

Finally, strong evidence was also presented to indicate that approval-dependent individuals show more double agreement. Since double agreers, in other situations, would have been classified erroneously as authoritarians, these results when taken collectively suggest that many bogus authoritarians are the result of ambiguous measuring instruments and perceived social pressures indigenous to the testing situation. We find ourselves in agreement with Peabody (1966) that the question is not so much whether there "ain't no such animal" (Samelson, 1964) as it is whether he is ensnared by F, Anti-Semitism, and Dogmatism scale traps.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BERG, I. A., & RAPAPORT, G. M. Response bias in an unstructured questionnaire. *Journal of Psychology,* 1954, **38**, 475–481.

CHAPMAN, L. J., & BOCK, R. D. Components of variance due to acquiescence and content in the F scale measure of authoritarianism. *Psychological Bulletin,* 1958, **55**, 328–333.

CHRISTIE, R., HAVEL, J., & SEIDENBERG, B. Is the F scale irreversible? *Journal of Abnormal and Social Psychology,* 1958, **56**, 143–159.

CROWNE, D. P., & MARLOWE, D. *The approval motive.* New York: Wiley, 1964.

MCBRIDE, L. K. A study of the relationship between item ambiguity, susceptibility to social influence, and the double agreement phenomenon. Unpublished master's thesis, University of Montana, 1966.

MIKLICH, D. R. Peabody's agreement response set measure: A reply to Samelson. *Psychological Reports,* 1966, **18**, 200–202.

PEABODY, D. Attitude content and agreement set in scales of authoritarianism, dogmatism, anti-Semitism, and economic conservatism. *Journal of Abnormal and Social Psychology,* 1961, **63**, 1–11.

PEABODY, D. Authoritarianism scales and response bias. *Psychological Bulletin,* 1966, **65**, 11–23.

ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychological Monographs,* 1956, **70**(18, Whole No. 425).

ROKEACH, M. The double agreement phenomenon: Three hypotheses. *Psychological Review,* 1963, **70**, 304–309.

SAMELSON, F. Agreement set and anticontent attitudes in the F scales: A reinterpretation. *Journal of Abnormal and Social Psychology,* 1964, **68**, 338–342.

STANLEY, G., & MARTIN, J. How sincere is the dogmatist? *Psychological Review,* 1964, **71**, 331–334.

(Received May 24, 1966)

Volume 6 June 1967 Number 2

Journal of Personality and Social Psychology

Editor: Daniel Katz, *University of Michigan*

Associate Editor: Robert B. Zajonc, *University of Michigan*

Robert P. Abelson
Yale University

Joseph B. Adelson
University of Michigan

Gordon W. Allport
Harvard University

Norman H. Anderson
University of California, San Diego

Elliot Aronson
University of Texas

Albert Bandura
Stanford University

Leonard Berkowitz
Cornell University

James Bieri
University of Texas

Jack W. Brehm
Duke University

Donald R. Brown
University of Michigan

Eugene Burnstein
University of Michigan

Arnold H. Buss
Rutgers University

Launor F. Carter
Systems Development Corporation

Rosalind D. Cartwright
University of Illinois
Medical Center, Chicago

Dwight W. Chapman, Jr.
Vassar College

Irvin L. Child
Yale University

Nickolas B. Cottrell
Kent State University

William Dember
University of Cincinnati

Morton Deutsch
Columbia University

Seymour Feshbach
University of Colorado

Donald W. Fiske
University of Chicago

Basil S. Georgopoulos
University of Michigan

Edwin P. Hollander
Tavistock Institute of Human Relations
London, England

Murray Horwitz
New York University

Herbert C. Kelman
University of Michigan

Daniel J. Levinson
Connecticut Mental Health Center, New Haven

Richard A. Littman
University of Oregon

Robert B. MacLeod
Cornell University

Melvin Manis
Veterans Administration Hospital
Ann Arbor, Michigan

David C. McClelland
Harvard University

Charles G. McClintock
University of California, Santa Barbara

William J. McGuire
Columbia University

Vincent Nowlis
University of Rochester

Harold L. Raush
University of Michigan

Milton E. Rosenbaum
University of Iowa

William A. Scott
University of Colorado

Janet Taylor Spence
University of Texas

Percy H. Tannenbaum
University of Wisconsin

Harry C. Triandis
University of Illinois

Joseph Veroff
University of Michigan

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Personality and Social Psychology* is devoted to basic research and theory in the broad field of social interaction and its determinants and effects. More specifically, it is concerned with the subareas of (*a*) social interaction and group process, (*b*) interpersonal perception and attitude change, (*c*) the psychological aspects of formal social systems and less structured collective phenomena (e.g., the public opinion process and mass movements), (*d*) the psychological analysis of cultural forms and products (e.g., language), (*e*) the socialization process both at the child and adult levels, (*f*) social motivation and personality dynamics, (*g*) the structure of personality, and (*h*) the relation of personality to group process and social systems. Animal studies as they relate to an understanding of social processes are definitely appropriate for the journal. Methodological articles and studies of the development of personality instruments are not excluded but are considered marginal unless presented as part of a substantive piece. Where existing journals already serve a given specialized interest within the broad domain of the field, articles appropriate to this area of specialization will have low priority for the *Journal of Personality and Social Psychology*.

The JOURNAL is issued monthly, three volumes per year. The subscription price is $30.00 annually (foreign, $30.50). Single issues are $3.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor-Elect:

Dr. William McGuire
Journal of Personality and Social Psychology
604 Mathematics Hall
Columbia University
New York, New York 10027

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the Editor or from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR	JOANNE BERGEN	HENRY S. SALMON	ROBERT D. GORMAN
Managing Editor	Technical Editor	Advertising and Business Manager	Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.

Copyright © 1967 by the American Psychological Association, Inc.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 119-125

EFFECTS OF PARENTAL DOMINANCE, WARMTH, AND CONFLICT ON IMITATION IN CHILDREN [1]

E. MAVIS HETHERINGTON AND GARY FRANKIE

University of Wisconsin

This study investigated the effects of parental dominance, warmth, and conflict on imitation of parents by boys and girls. Parental warmth and dominance were found to be salient variables in identification; however, parental dominance was more important for imitation by boys while maternal warmth was more effective with girls. Support was found for identification with the aggressor under the conditions of a high-conflict home where both parents were low in warmth.

Most theories of identification agree that identification is based on a process or processes whereby the child, through imitation, modeling, or introjection acquires traits, characteristics, and values similar to the parents. Although there is agreement in the defining characteristics of identification, the various theories diverge in their emphases on the relative importance of different motivational and learning conditions leading to identification. Three variables have frequently been hypothesized by the different theories as affecting identification; namely, parental power, parental warmth, and parental aggression.

Parsons (1955) has emphasized the importance of total parental power in the development of identification. According to Parsons, the child identifies with the parent because he determines or mediates both the rewards and punishments the child receives. Several studies (Hetherington, 1965a; Mussen & Distler, 1959, 1960) do indeed suggest that parental power or dominance influences sex typing in boys. The Hetherington (1965a) study also indicated that parental dominance had little effect on sex typing in girls, although paternal dominance increased father-daughter similarity on non-sex-typed traits without interfering with mother-daughter similarity.

Learning theorists focus on anaclitic identification and the effects of warmth and nurturance in the development of identification. Considerable evidence has been accumulated indicating that identification and appropriate sex-role typing are facilitated for both-sex children by warmth in the same-sex parent (Helper, 1955; Mussen & Distler, 1959, 1960; Mussen & Rutherford, 1963; Payne & Mussen, 1956; Sears, 1953).

Psychoanalytic theorists have emphasized the role of fear of punishment and defensive identification with a threatening model. Most of the support for this position has been in clinical case studies, anecdotal evidence, or naturalistic observations such as the German concentration camp studies (Bettelheim, 1943). The Mussen and Distler (1959, 1960) studies did offer some evidence that highly masculine boys perceived their fathers as more punitive as well as more nurturant than feminine boys. No relationship was found between punitiveness and sex typing in girls in a subsequent study by Mussen and Rutherford (1963).

Bandura (1962) has suggested that identification and imitation are synonymous since both encompass the tendency for a person to match the behavior, attitudes, or emotional reactions exhibited by models. If this is true, then the same variables thought to be significant in identification should be salient in the child's imitation of the parent. Experimental studies have found that children imitate a powerful model (Bandura, Ross, & Ross, 1963) or a nurturant model (Bandura & Huston, 1961). Two studies using parents as models have also found that both boys and girls imitate the dominant parent (Hetherington, 1965a), and that maternal nurturance increases imitation by daughters (Mussen &

---

[1] This study was supported by the Research Committee of the Graduate School of the University of Wisconsin with funds provided by the Wisconsin Alumni Research Foundation.

Parker, 1965). Although there is evidence that children imitate aggressive behavior in others, there is only limited evidence that they emulate the behavior of an aggressive model when the aggression is directed toward themselves. A recent study by Mischel and Grusec (1966) has found that children rehearsed aggressive behaviors directed at them more when the model was high in control, and transmitted the aversive behaviors more when the model was high in rewardingness.

Perhaps one reason why evidence for identification with the aggressor is at best suggestive is that it can only be found in certain restricted circumstances. Sarnoff (1951) has suggested that three conditions are essential in producing defensive identification: a hostile individual who directs his aggression toward another person, a victim who is dependent upon the aggressor, and a situation involving stresses and limitations which prevent the victim from escaping the hostile behavior of the aggressor. On the basis of clinical and sociological observations, and Sarnoff's criteria for identification with the aggressor, defensive identification with a hostile dominant parent seems most likely to occur in a stressful home in which both parents are lacking in warmth. Such a home situation would offer the child no escape by seeking a closer relationship with a warm nondominant parent. A stressful, conflictual family relationship should add to the child's feeling of helplessness and increase his tendency toward defensive identification. The present study attempted to test this hypothesis.

A second purpose of the present study was to investigate further the effects of the different variables postulated to affect identification. There seems to be ample evidence that warmth and power do affect identification; a necessary step would appear to be to investigate the interactions of these variables and to find if there are situations under which warmth or dominance are particularly influential. For example, there has been some suggestion in the literature that parental power is more important in the identification of boys than of girls and that warmth is more salient for girls (Hetherington, 1965b; Mussen & Rutherford, 1963).

To test these hypotheses the present study investigated the effects of parental warmth, dominance, and conflict on imitation of parents by boys and girls.

METHOD

Subjects

Subjects were 80 male and 80 female nursery-school and kindergarten children and their parents randomly selected from a large pool of 310 families in which the parents had already taken the Structured Family Interaction Task. These families lived in small Wisconsin towns and most would probably be classified as lower middle class. The ages of the children ranged from 4 years, 4 months to 6 years, 5 months. Half of the subjects were from high-conflict homes and half were from low-conflict homes. Within each conflict group half of the subjects were from mother-dominant homes and half from father-dominant homes. Groups were further subdivided on the basis of all possible mother-father warmth combinations. Thus within each conflict-dominance group there were four warmth combinations: mother high–father low, mother low–father high, mother high–father high, and mother low–father low.

The measures of dominance, conflict, and maternal and paternal warmth were not significantly correlated, although there was an insignificant trend for parental warmth to be associated with low conflict. The number of subjects within each cell in the total populations sampled was approximately equal. These ranged from $N = 13$ in the father-dominant, low-conflict, both-parents-warm, boys group to $N = 9$ in the father-dominant, high-conflict, both-parents-warm, girls group.

Experimental Design

The study utilized a $2 \times 2 \times 2 \times 2 \times 4$ mixed factorial design involving parental dominance, conflict, sex of subject, parent imitated, and parental warmth combination. Each individual subject had two scores, one for imitation of his mother and one for imitation of his father.

Parental Measures

The parental measures of warmth, hostility, conflict, and dominance were obtained from a structured family-interaction task adapted from a procedure developed by Farina (1960). Each parent was seen individually in a quiet room in his own home. He was read 12 hypothetical problem situations involving child behavior and asked how he would handle them when he was by himself. The instructions were as follows:

> We are interested in knowing how a father/mother handles situations that come up when his wife/husband is not around. I'm going to read some situations that ——— might or might not really have been involved in. Imagine that this situation has arisen and that your wife/husband is not around and you must handle the problem yourself.

Both parents were then brought together and asked to arrive at a compatible solution on handling these children's problems.

The instructions for the joint interaction sessions were as follows:

> You have talked about how you would handle these various situations if you were alone; now I would like to go through these situations again, and have you discuss them and come to some agreement as to how you would handle the problem if you were both there. Imagine the situation arising, you are both at home and must deal with the situation. I want you to continue the discussion until you can come to some agreement on how you would handle the problem if you were together, then say "Agreed" and we will go on to the next situation.

The discussion of each problem continued until both parents said the terminating signal, "agreed." The experimenter participated only minimally in the discussion in order to clarify scoring responses. All interviews were tape-recorded and scored later.

In the previous studies, for which the interaction task was initially run (Hetherington, 1965a, 1965b), the results of all 12 situations were used; however, since many parents complained of fatigue and became restless during the last part of the joint interaction sessions only the responses on the first 7 situations were utilized in this study.

The seven situations were as follows:

1. Your son/daughter loses his/her temper while playing with a toy and intentionally breaks it.

2. You have friends over in the evening. Your son/daughter keeps getting out of bed to see what's going on.

3. Your son/daughter has a friend over to play. The friend wants to play with one of ________'s favorite toys but ________ won't let him.

4. A neighbor calls up and complains that your son/daughter has been throwing rocks at her child.

5. You have gone out of your way to buy something nice for your child and then he/she throws it aside and says he/she doesn't like it.

6. Your child has been asked several times to tidy up his/her room. You find his/her room still a mess and him/her watching TV.

7. You have taken your son/daughter out to dinner in a restaurant as a special treat. He/she is behaving in a generally noisy, ill-mannered way although you have warned him/her to quiet down.

Parental dominance measure. The index for parental dominance was comprised of five of the measures previously used by Farina (1960), which were: speaks first, speaks last, passive acceptance of spouse's solution, percent of total speaking time, and the amount of yielding from original individual solution to the joint solution. Total scores on each of these measures were used to classify the index as indicating mother or father dominance. If three or more indexes indicated paternal dominance, the family was classified as father dominant; if three or more indicated maternal dominance, the family was classified as mother dominant.

Parental conflict measure. The parental conflict measures were the same as those used by Farina (1960): total time spoken, disagreements and aggressions, interruptions, frequency of simultaneous speech, and failure to agree. Total time spoken was included as a measure of parental conflict, since it was assumed that the greater the conflict and disagreement the longer it would take to resolve differences and come to a mutually acceptable solution on handling problems. These scores were converted to z scores and combined into a single conflict index. Families which scored above the group mean in conflict were classified as high-conflict homes; those below the mean were classified as low-conflict homes.

Warmth-Hostility measure. The mother and father were separately rated on a 6-point warmth-hostility scale ranging from 1—extremely warm, nurturant, and affectionate; clearly proud of the child, concerned with and enjoys the child as a person; understanding and empathic—to 6—marked hostility, anger, and punitiveness toward the child; little sympathy or attempt to understand the child's behavior; always interprets the child's behavior in the worst light. Ratings of parental warmth were done on the basis of both the individual sessions and the joint interaction session. A parent scoring below the group mean for his sex was classified as high in warmth; those above the mean were classified as low in warmth.

All the measures on the Structured Family Interaction Task which had previously exhibited any unreliability were rated by two judges. These included the yielding measure of the dominance scale, the warmth-hostility scale, and the conflict measures. Interjudge reliability was .94 for the yielding measure, .81 for the warmth-hostility ratings, and mean interjudge reliability on the conflict measures was .86.

Imitation Task

Each child was run on an imitation task where he watched each parent alternately perform for four trials in a free-play situation. Each parent was preinstructed, given a practice session without the child present, and given a small inconspicuous card to carry summarizing his role in the imitation task. Following the second, fourth, sixth, and eighth trials the child was given a 5-minute session in the playroom. The parental behaviors involved postural, motor, and verbal responses associated with playing with a group of toys and games previously determined to be of equal interest to male and female children. An attempt was made to use activities and behaviors which were appropriate for adults rather than some of the bizarre behaviors used in previous studies. Thus the activities involved were some in which an adult might participate without appearing too ludicrous, such as golf putting, shooting rubber darts at a target, a game of ball throwing at a target which automatically ejected the ball, etc. The parental behaviors involved such things as making

TABLE 1

Analysis of Variance of the Imitation Scores

Source	*df*	*MS*	*F*
Conflict (C)	1	116.40	2.85*
Dominance (D)	1	1.95	<1.00
Sex of *S* (S)	1	2.28	<1.00
Warmth (W)	3	600.03	14.71***
C × D	1	14.03	<1.00
C × S	1	3.00	<1.00
C × W	3	10.61	<1.00
D × S	1	114.00	2.80*
D × W	3	32.19	<1.00
S × W	3	59.91	1.47
C × D × S	1	.03	<1.00
C × D × W	3	2.80	<1.00
C × S × W	3	3.55	<1.00
D × S × W	3	40.35	<1.00
C × D × S × W	3	13.83	<1.00
Error (a)	128	40.78	
Parent imitation (P)	1	717.00	19.77***
C × P	1	3.83	<1.00
D × P	1	4096.95	112.99***
S × P	1	1688.20	46.56***
W × P	3	1048.75	28.92***
C × D × P	1	141.79	3.91*
C × S × P	1	8.78	<1.00
C × W × P	3	4.56	<1.00
D × S × P	1	1762.51	48.61***
D × W × P	3	11.79	<1.00
S × W × P	3	61.64	1.70
C × D × S × P	1	2.27	<1.00
C × S × W × P	3	7.56	<1.00
C × D × W × P	3	53.06	1.46
D × S × W × P	3	45.35	1.25
C × D × S × P × W	3	16.41	<1.00
Error (b)	128	36.26	

* $p < .10$.
*** $p < .01$.

predetermined distinctive comments following success or failure, always selecting a toy of a given color and saying, "———— is my lucky color," squatting and lining up golf shots, pulling up a chair, sitting sideways and shooting with two hands in the dart game, and so on.[2] The set of imitative responses assigned to a given parent and the parent performing first were randomly determined. Parents were always absent from the room during the child's test series.

A male and female observer checked the imitative responses on a response check list; perfect interrater agreement was found in the scoring of 95% of the specific imitative responses. The imitation scores were obtained by summing the frequency of responses the child made which were similar to those of a given parent.

Results and Discussion

The basic analysis of variance for the imitation scores is presented in Table 1. The significant main effect of parent imitation and the significant Sex of Subject × Parent Imitation interaction indicate that although mothers were imitated more than fathers, boys imitated the father more than the mother (M = 16.82, 15.22, respectively, $p < .10$),[3] while girls imitated the mother more than the father ($M = 19.65$, 12.06, respectively, $p < .05$). The significant Dominance × Parent Imitation and Warmth × Parent Imitation interactions support the previous findings that parental warmth and dominance are important factors in identification. The means for the Dominance × Parent Imitation interaction show that the dominant parent was imitated more; in a mother-dominant home the means for imitation of the mother and father were 20.94 and 10.78, respectively ($p < .05$), and in a father-dominant home means for imitation of the mother and father were 13.94 and 18.10, respectively ($p < .05$). In looking at the means for the Warmth × Parent Imitation interaction, it is obvious that a parent high in warmth is imitated more than a parent low in warmth. Mean imitation of the mother under the various warmth conditions was as follows: mother high–father high, 21.25; mother low–father high, 12.85; mother high–father low, 23.27; and mother low–father low, 12.40 (critical difference = 2.73, $p < .05$). It can be seen that a mother high in warmth was imitated more than one low in warmth, regardless of the level of paternal warmth. The same type of results were also obtained for mean imitation of the father under the various warmth conditions: father high–mother high, 16.72; father low–mother high, 10.92; father high–mother low, 17.35; and father low–mother low, 12.77 (critical difference = 2.73, $p < .05$). Also significant ($p < .05$) is the difference between imitation of a highly warm mother and a highly warm father (i.e., comparing imitation of the mother in the warmth conditions of mother high–father high, and mother high–father low versus imitation of the father in the conditions of father high–mother high and father high–mother low). This finding suggests that maternal warmth facilitates imitation of the mother more than paternal warmth facilitates imitation of the father.

---

[2] Complete details on the imitation procedures may be obtained by writing to the senior author.

[3] Probability values are based on two-tailed t tests, unless otherwise noted.

TABLE 2

MEANS FOR THE DOMINANCE × SEX × PARENT IMITATED INTERACTION

Parent imitated	Mother dominant		Father dominant	
	Boys	Girls	Boys	Girls
Mother	20.47	21.40	9.98	17.90
Father	10.22	11.35	23.42	12.78

Note.—Critical difference (2-tailed t test, $p < .05$) = 2.67 for comparison of mother versus father imitation within a column, and 2.73 for comparisons across columns.

However, these general findings must be qualified when we look at the means for the significant Dominance × Sex × Parent Imitated interaction presented in Table 2. Under mother dominance both boys and girls imitated the mothers more. Under father dominance, however, boys imitated the father more while girls continued to imitate the mother. This suggests that parental dominance has a more important effect on the identification of boys than of girls. The age range of the subjects in this study, 3–5, is considered a particularly important transition period in identification for boys where identification must shift from the mother to the father. Paternal dominance may play an extremely salient role in facilitating this shift for boys. It might also be suggested that a dominant father offers a more appropriate role model for boys than does a nondominant father.

For the girls, who do not need to shift their identification, maternal warmth seems to be the most salient variable in imitation. When the warmth-combinations variable is broken down into the orthogonal factors of mother and father warmth, the interaction of these factors with sex of subject (presented in Table 3) shows that maternal warmth interacts significantly with sex of subject while paternal warmth facilitates imitation to an equal degree in boys and girls. The means for the significant Maternal Warmth × Sex of Subject interaction are presented in Table 4 and show that maternal warmth affects the girl's imitation more than it does the boy's.

TABLE 3

ANALYSIS OF VARIANCE FOR SEX OF SUBJECT × WARMTH EFFECTS INTERACTIONS

Source	df	MS	F
S × Warmth	3		
S × MW	1	161.03	3.95**
S × FW	1	1.13	<1.00
S × MW × FW	1	17.58	<1.00

Note.—S = sex, MW = mother warmth, FW = father warmth.
** $p < .05$.

TABLE 4

MEANS FOR THE SIGNIFICANT SEX OF SUBJECT × MATERNAL WARMTH INTERACTION

Sex	Maternal warmth	
	High	Low
Girls	18.66	13.05
Boys	17.41	14.64

The means for the marginally significant ($p < .06$) Conflict × Dominance × Parent Imitated interaction are presented in Table 5. The pattern of differences in this table suggests that in a stressful home situation having high conflict there is more imitation of the dominant parent than is found in a home having low conflict. This appears to be particularly true if the mother is the dominant parent.

In order to permit clearer evaluation of the interaction of specific variables, and in order to assess the possibility of defensive identification occurring under the previously hypothesized conditions of Sarnoff's (1951) study, the means for all subgroups are presented in Table 6. The means essential for the defensive identification analysis are outlined.

It was assumed that the conditions most likely to lead to identification with the aggressor would be a home in which there is a

TABLE 5

MEANS FOR THE CONFLICT × DOMINANCE × PARENT IMITATED INTERACTION

Conflict	Mother dominant		Father dominant	
	Parent imitated		Parent imitated	
	Mother	Father	Mother	Father
High	22.52	10.82	13.77	19.05
Low	19.35	10.75	14.10	17.15

Note.—Critical difference = 2.73, 2-tailed t test, $p < .05$.

TABLE 6

MEAN IMITATION SCORES

Homes	Parent imitated	Warmth combinations							
		Mother low–Father low		Mother high–Father low		Mother low–Father high		Mother high–Father high	
		Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Mother dominant High conflict	Mother	20.6	20.2	28.4	27.6	15.0	16.6	25.6	26.2
	Father	8.8	7.8	7.2	9.0	12.4	15.6	12.0	13.8
Low conflict	Mother	13.2	14.2	25.6	27.2	15.0	15.0	20.4	24.2
	Father	10.5	10.0	6.4	8.6	12.0	14.0	12.8	12.0
Father dominant High conflict	Mother	6.8	7.6	13.0	25.4	9.0	11.8	11.2	25.4
	Father	22.5	15.6	19.8	10.4	27.8	14.6	27.8	14.2
Low conflict	Mother	7.0	9.6	13.6	25.4	7.4	13.0	11.8	25.0
	Father	20.6	7.0	16.4	9.6	26.2	16.2	26.6	14.6

Note.—Means for evaluating defensive identification under Sarnoff's (1951) criteria are outlined. Critical difference = 6.3, 1-tailed *t* test, $p < .05$, for comparing mother imitation versus father imitation at a given level of all other variables. For other comparisons critical difference = 6.5, 1-tailed *t*, $p < .05$.

high conflict and in which both parents were low in warmth. In such a situation there would be no warm supportive parent to whom to turn for succor so that the child might attempt to minimize his insecurity by identifying with a powerful, punitive model. Under the less stressful situation of a low-conflict home or a home in which there is some protection from a warm parent, there should be less tendency to imitate a hostile aggressive parent. Since these assumptions produce directional hypotheses, one-tailed *t* tests were used in testing the differences in Table 6.

It can be seen that under high conflict, with both parents low in warmth, there is indeed a significant tendency for both boys and girls to imitate the dominant parent regardless of sex of the parent. If either the nondominant parent is warm or conflict is reduced, there is a trend toward less imitation of the aggressive dominant parent, so that the dominant parent is not imitated significantly more than the nondominant one. This trend does not hold in the case of boys with dominant fathers, however. The boy's tendency to imitate a dominant father overrides the effects of variations in conflict and warmth. In contrast, maternal warmth appears to be particularly salient for girls. Even under conditions of high conflict and paternal dominance there is marked imitation of warm mothers by daughters.

In summary, the results seem to be congruent with those of past studies which have found that both parental warmth and power are important in the identification of girls, and paternal dominance is important in the identification of boys. This is in agreement with the earlier findings of Mussen and Rutherford (1963) and Mussen and Distler (1959, 1960). Some support was found for identification with the aggressor under very restricted conditions involving high stress and low warmth in both parents, which might be assumed to result in a sense of extreme helplessness on the part of the child.

REFERENCES

BANDURA, A. Social learning through imitation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1962*. Lincoln: University of Nebraska Press, 1962. Pp. 211–269.

BANDURA, A., & HUSTON, A. C. Identification as a process of incidental learning. *Journal of Abnormal and Social Psychology*, 1961, **63**, 311–318.

BANDURA, A., ROSS, D., & ROSS, S. A. A comparative test of the status envy, social power, and the secondary-reinforcement theories of identification learning. *Journal of Abnormal and Social Psychology*, 1963, **67**, 527–534.

BETTELHEIM, B. Individual and mass behavior in extreme situations. *Journal of Abnormal and Social Psychology,* 1943, **38**, 417–452.

FARINA, A. Patterns of role dominance and conflict in parents of schizophrenic patients. *Journal of Abnormal and Social Psychology,* 1960, **61**, 31–38.

HELPER, M. M. Learning theory and the self-concept. *Journal of Abnormal and Social Psychology,* 1955, **51**, 184–194.

HETHERINGTON, E. M. A developmental study of the effects of sex of the dominant parent on sex-role preference, identification, and imitation in children. *Journal of Personality and Social Psychology,* 1965, **2**, 188–194. (a)

HETHERINGTON, E. M. The effects of parental dominance on imitation of sex typed behaviors. Unpublished manuscript, University of Wisconsin, 1965. (b)

MISCHEL, W., & GRUSEC, JOAN. Determinants of the rehearsal and transmission of neutral and aversive behaviors. *Journal of Personality and Social Psychology,* 1966, **3**, 197–205.

MUSSEN, P., & DISTLER, L. Masculinity, identification and father-son relationships. *Journal of Abnormal and Social Psychology,* 1959, **59**, 350–356.

MUSSEN, P., & DISTLER, L. Child rearing antecedents of masculine identification in kindergarten boys. *Child Development,* 1960, **31**, 89–100.

MUSSEN, P., & RUTHERFORD, E. Parent-child relations and parental personality in relation to young children's sex-role preferences. *Child Development,* 1963, **34**, 589–607.

MUSSEN, P. H., & PARKER, A. L. Mother nurturance and girls' incidental imitative learning. *Journal of Personality and Social Psychology,* 1965, **2**, 94–97.

PARSONS, T. Family structure and the socialization of the child. In T. Parsons & R. F. Bales (Eds.), *Family socialization and interaction process.* Glencoe, Ill.: Free Press, 1955. Pp. 35–131.

PAYNE, D. E., & MUSSEN, P. H. Parent-child relations and father identification among adolescent boys. *Journal of Abnormal and Social Psychology,* 1956, **52**, 358–362.

SARNOFF, I. Identification with the aggressor: Some personality correlates of anti-semitism among Jews. *Journal of Personality,* 1951, **20**, 199–218.

SEARS, P. S. Child-rearing factors related to playing of sex-typed roles. *American Psychologist,* 1953, **8**, 431. (Abstract)

(Received August 11, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 126–131

ENERGIZING (DYNAMOGENIC) EFFECT OF COGNITIVE DISSONANCE ON TASK PERFORMANCE[1]

CAROLINE K. WATERMAN AND EDWARD S. KATKIN

State University of New York at Buffalo

Drives, in addition to motivating behaviors which will reduce them, are thought to have a general energizing effect on whatever response tendencies exist in a given situation. This experiment was designed to show that cognitive dissonance energizes behaviors that are not dissonance reducing. Ss in a dissonant group wrote attitude-discrepant essays while Ss in a consonant group wrote attitude-consistent essays. Under the guise of a separate experiment, all Ss performed a simple and a complex symbol-substitution task. As predicted, dissonance facilitated performance on the simple task. Contrary to expectation, dissonance also facilitated performance on the complex task.

Cognitive dissonance has been conceptualized by Festinger (1957) as a "motivating factor [for which] one can substitute other notions similar in nature such as 'hunger' [p. 3]." Support for Festinger's assertion that dissonance possesses drive properties has been derived from the frequent demonstrations, in a variety of free-choice and forced-compliance situations, that its reduction apparently is reinforcing. On the basis of such demonstrations Brehm and Cohen (1962), for example, have stated that dissonance is a drive because "its arousal always elicits behavior that is . . . directed to returning the organism to a state in which the drive is at a lower level of arousal [p. 224]."

A drive, however, in addition to motivating behaviors which will reduce it, is thought to have a general energizing (dynamogenic) effect on whatever response tendencies exist in a given situation, whether or not those responses can reduce that drive (see Farber, 1954). Although the energizing effect of a drive on behavior which does not reduce that drive has been demonstrated clearly for primary drives such as hunger (Amsel, 1950), thirst (Webb, 1949), and pain (Ullman, 1951) as well as for the secondary drive of anxiety (Spence, 1958; Spence & Spence, 1966; Taylor, 1956), it remains a moot question whether this effect can be demonstrated for cognitive dissonance. The purpose of the present experiment, therefore, was to examine the energizing properties of cognitive dissonance.

Dissonance was aroused in one group of subjects by asking them to write attitude-discrepant essays, while consonant attitudes were established in another group by asking them to write attitude-consistent essays. Then a comparison was made of the performance of these two groups on independent tasks. Although questions have been raised about the utility of anonymous essay writing as a dissonance inducer (Carlsmith, Collins, & Helmreich, 1966), an extensive body of research (see Brehm & Cohen, 1962; Kelman, 1953; Rabbie, Brehm, & Cohen, 1959) has demonstrated that the writing of counterattitudinal essays results in cognitive reevaluation assumed to result from a state of dissonance.[2] In the present experiment, freshman and sophomore students were asked to write essays either favoring (dissonant condition) or opposing (consonant condition) the proposal that parking privileges on campus be restricted to seniors, graduate students, faculty, and staff. This issue was chosen for two reasons: first, the overcrowded parking situation had received considerable coverage in the campus newspaper for several weeks; second, adoption of this proposal would have impor-

---

[1] This investigation was supported in part by Research Grant MH 11989 from the National Institute of Mental Health, United States Public Health Service, awarded to the second author.

[2] In some of these studies the investigators do not describe whether the essay writing was anonymous. In others, they describe situations in which the essay was signed, or the identity of the subjects was made known to the experimenter in some other fashion. In this study, subjects were *not* anonymous.

tant behavioral implications for lowerclassmen who used the campus parking facilities.

In order to evaluate the behavioral consequents of dissonance induction with a task which was presumed to be independent from the specific dissonance-producing situation, symbol-substitution tasks were used, since neither success nor failure on such tasks could be regarded as a means of reducing the dissonance aroused by writing discrepant essays. However, when assessing the relationship between drive level and task performance, special attention must be paid to the nature of the tasks employed. According to Hullian theory, drive interacts multiplicatively with all responses in a subject's hierarchy; thus, when a task elicits few competing responses, and their strength is low, the effect of high drive level is to facilitate and enhance task performance. On the other hand, when a task elicits many competing response tendencies, drive level energizes them all, and thus is found often to be inversely related to task performance (Farber & Spence, 1953; Montague, 1953; Ramond, 1953; Spence, Farber, & McFann, 1956; Taylor & Spence, 1952). Therefore, the present experiment utilized two symbol-substitution tasks designed to differ in the number and strength of competing response tendencies which they would evoke. The following hypotheses were formulated: (*a*) On a task with few competing responses (simple task), subjects who have written attitude-discrepant essays will complete more symbols and demonstrate better recall of them than subjects who have written attitude-consonant essays; (*b*) on a task with many competing responses (complex task) subjects who have had dissonance induced will complete fewer symbols and demonstrate poorer recall of them than consonant subjects.

METHOD

Subjects

The subjects were 34 male freshmen and sophomores who required the use of campus parking facilities at least once a week. They had volunteered for a "survey on the parking problem on campus" and received one credit toward fulfilling a course requirement for participation in experiments.

The subjects were assigned randomly either to the dissonant or consonant group with the only restriction on randomness being the establishment of equal size groups. During the course of the experiment two subjects were dropped from the dissonant group (one for favoring the proposal strongly, and one for verbalizing spontaneously his knowledge of the purpose of the experiment).

Procedure

When the subjects arrived at the lab, they were introduced to a female experimenter who seated them at a table and said:

> As you know, you signed up for 40 minutes worth of participation. You're probably aware that the subject pool is not large enough to meet all the research needs of the department. Therefore, you will actually be participating in two experiments today. This one will take only 25 minutes of your time. After that you will participate for 15 minutes in an experiment being run by one of Dr. ________'s assistants in room #_____.

Then the subjects were told that "an interdepartmental faculty committee has asked the psychology department to conduct a survey for them about attitudes toward the parking problem on campus." Subjects were asked if they had heard anything about the survey. None of them reported having any prior information about the survey, but many expressed their familiarity with the newspaper coverage of the problem. The experimenter told them that "one of the solutions that they have proposed for the overcrowding problem is the limiting of parking privileges on campus to seniors, graduate students, faculty and staff," and that the survey would be concerned with that proposal.

The subjects were then asked to fill out a five-item questionnaire designed to evaluate the extent of difficulty which they experienced in parking on campus, and how often they required campus parking facilities. To determine initial attitudes toward the proposal one of the questions was, "How do you feel about a proposal to limit parking privileges on campus to seniors, graduate students, faculty and staff?" Each question was answered on a 7-point Likert scale; the question concerning attitude toward the proposal was answered on a scale ranging from "strongly favor" (1) to "strongly oppose" (7).

After subjects answered the questionnaire the experimenter said:

> The faculty committee is interested in getting arguments on both sides of this issue. It has been shown that one of the best ways to get relevan arguments on both sides of an issue is to as people to write essays favoring only one side o the issue. We think we know pretty much how you feel about the restriction of on-campus parking privileges. What we really need now are som essays favoring (opposing) this plan. I woul like you to write the strongest, most forcefu most creative, and thoughtful essay you can, unequivocally favoring (opposing) this plan. You wi have ten minutes to write this. Remember, we a more interested in the arguments you present tha in writing style.

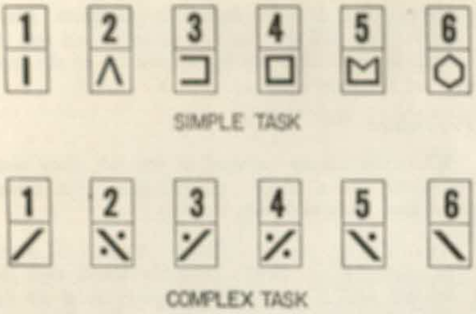


FIG. 1. Symbols used in the simple and complex tasks.

These instructions are essentially the same as those reported by Cohen (see Brehm & Cohen, 1962, pp. 73–78).

The subjects were then given an examination booklet in which to write their essays, and were asked to put their names on the cover. They were handed a 3 × 5-inch white index card containing the title of the essay to be written: "Why I believe (do not believe) that parking privileges should be restricted to seniors, graduate students, faculty, and staff." If a subject expressed hesitancy about writing an attitude-discrepant essay, the experimenter assured him that he was free to leave if he desired. No subject chose to leave the experiment.

While subjects were writing the essays the experimenter left the room. She returned 10 minutes later, and if a subject had not finished writing his essay he was permitted to finish the sentence he was writing, and thanked for participating.[3] Then the subjects were directed to the second experimenter in another room. The second experimenter (male), who was completely unaware of which essay the subject had written, explained that he was working for Dr. ———— "who is interested in developing some alternate forms for one of the scales of the Wechsler Adult Intelligence Scale" (WAIS). He continued:

> As you may know, it is an individual intelligence test, composed of eleven subscales, one of which is a symbol substitution task. The symbol substitution task that is currently being used has come under criticism recently. What we are trying to do is to see if a more adequate task can be constructed, and we are currently testing two possible forms which I would like you to take.

The instructions for completing the task were identical to those recommended by Wechsler (1955) for administering the symbol substitution subtest of the WAIS. The symbols used in the present investigation are presented in Figure 1; the format of presentation was identical with that of the WAIS.[4] By random determination half of the subjects took the simple task first, followed by the complex task, and the other half took the complex task first, followed by the simple task. In this manner four groups of subjects were established: dissonant-simple first, dissonant-complex first, consonant-simple first, and consonant-complex first. All subjects were given a 90-second time limit to perform each task. The "simple" task was constructed in such a way that the number of lines involved in the production of any given symbol was the same as the number associated with that symbol, and each symbol was a uniquely different configuration. Thus, competing response tendencies were presumed to be low. The "complex" task was designed to have high intersymbol similarity, the symbols differing only with regard to the direction of the lines and the number and placement of surrounding dots. Following the completion of both tasks, the subjects were asked to write down as many symbols as they could remember next to the appropriate number. For this purpose, a sheet of paper with two sets of the numbers 1 through 6 on it was provided. Thus, a subject's recall for the symbols could be evaluated as well as his performance on the tasks.

To summarize, the experiment consisted of a 2 × 2 × 2 design in which dissonance and order of presentation were between-subjects variables and type of task was a within-subjects variable (Type III, Lindquist, 1953).

RESULTS

Before the symbol-completion and recall data were subjected to analysis, the responses to the preliminary questionnaire were evaluated in order to determine whether the groups differed in attitude toward the proposal at the outset of the experiment. Analyses of variance of the means of the four experimental groups revealed no significant differences among them on any of the five questions (F ratios ranged from .39 to 1.63; $df = 3/28$). Thus, the evaluation of the substitution and recall data which follows can be assumed to be independent of any preexisting differences in attitude among the groups.

TABLE 1

MEAN NUMBER OF SYMBOLS COMPLETED CORRECTLY ON SIMPLE AND COMPLEX TASKS FOR SUBJECTS IN THE FOUR EXPERIMENTAL CONDITIONS

Group	Simple task 1st		Complex task 1st	
	Simple	Complex	Simple	Complex
Dissonant	56.50	58.75	54.88	56.63
Consonant	51.25	51.38	49.88	53.50

---

[3] No differences were found between subjects in the dissonant and consonant conditions regarding the length of the essays written or the amount of time needed for completion of them.

An overall mean of 6.72 was obtained on the question concerning subjects' feelings toward the proposal, indicating that the proposal was indeed discrepant with attitudes of all subjects.

Symbols Completed

The mean number of symbols completed correctly for subjects in each of the four experimental groups is presented in Table 1, in which it may be noted that the performance of the subjects in the dissonant group was superior to that of the subjects in the consonant group, irrespective of task difficulty or order of presentation. Contrary to expectation, dissonant subjects did not perform more poorly than consonant subjects on the complex task. Analysis of variance of these data confirmed these observations, yielding a significant main effect of the dissonance manipulation ($F = 4.82$; $df = 1/28$; $p < .05$), and a main effect of task difficulty approaching significance ($F = 3.41$; $df = 1/28$; $.05 < p < .10$). No other significant main effects, nor any significant interaction effects were obtained. Thus, the hypothesis that cognitive dissonance would enhance performance on a simple task was confirmed, but the notion that the effects would be reversed on a complex task was not confirmed.

Recall

The mean number of symbols recalled correctly is presented in Table 2, in which a number of trends may be discerned. First, the dissonant subjects showed better recall than the consonant subjects, irrespective of task difficulty and order of presentation. However, analysis of the data indicated that this finding fell just short of traditional levels of significance ($F = 3.13$; $df = 1/28$; $.05 < p < .10$). Table 2 also indicates that all subjects demonstrated better recall for the simple task than for the complex task, and that the discrepancy between recall of the tasks seemed to be greater when the complex task was presented first. Analysis of variance confirmed both of the observations, yielding a significant main effect of task difficulty ($F = 84.87$; $df = 1/28$; $p < .001$), and a significant Task Difficulty × Order interaction ($F = 8.13$; $df = 1/28$; $p < .01$). In addition, a main effect of order was obtained ($F = 8.26$; $df = 1/28$; $p < .01$); however, observation of the data indicates that this main effect is attributable entirely to the fact that all subjects demonstrated better recall for the complex task when it was presented second rather than first. It is obvious that order had no effect whatever on recall of the simple task. Therefore, it appears that the main effect of order is an artifact of the Order × Task Difficulty interaction.

TABLE 2

MEAN NUMBER OF SYMBOLS RECALLED CORRECTLY ON SIMPLE AND COMPLEX TASKS FOR SUBJECTS IN THE FOUR EXPERIMENTAL GROUPS

Group	Simple task 1st		Complex task 1st	
	Simple	Complex	Simple	Complex
Dissonant	6.00	4.50	6.00	2.50
Consonant	5.63	3.50	5.63	2.25

Thus the data on recall also are consistent with the hypothesis that cognitive dissonance enhances performance on a simple task, and inconsistent with the hypothesized reversal on the complex task.

DISCUSSION

In the present study it was found that cognitive dissonance enhanced task performance on both simple and complex tasks. Further, it was found that all subjects, irrespective of dissonance or order of presentation of the tasks, tended to complete more complex symbols than simple ones, but clearly recalled more simple symbols than complex ones. Finally it was discovered that task difficulty and order of presentation interactively affected recall, but had no apparent effect on number of symbols completed.

The significant main effect of the dissonance manipulation on symbol substitution and the near significant effect on recall support the notion that cognitive dissonance has facilitative effects on behavior, similar to the energizing effects which have been demonstrated for other drives. It might be argued that the findings for completion and recall are not entirely independent, since subjects in the dissonant group completed more symbols and thus had greater opportunity to become familiar with them, thereby facilitating recall. However, this possibility appears unlikely, since Pearson product-moment correlations between number of symbols completed and number recalled were not significant on either the simple ($r = .11$) or the complex ($r = .13$) task.

Although the main effects described above may be interpreted as partial support for the notion that dissonance has energizing properties, the finding that dissonant subjects tended to complete and recall more symbols than consonant subjects on *both* tasks was not consistent with the hypothesis that dissonance would disrupt performance on the complex task. Two possible explanations for the failure to obtain a Task Difficulty × Dissonance interaction on the symbol-completion data may be formulated.

First, Spence and Spence (1966) have suggested that on complex tasks, a "crossover" effect may be obtained in which high-drive subjects initially show inferior performance, but as they gain experience at the task and develop mastery the correct response becomes differentially energized, competing responses drop out, and their performance becomes superior to that of the low-drive group. It is possible that in the present study the dissonant subjects performed more poorly on the complex task during, say, the first 30 or 40 seconds of the task, then began to demonstrate task superiority, resulting in better performance.

The second possible reason for the failure to obtain the predicted Dissonance × Task Difficulty interaction on the completion data may be simply that the complex task was not significantly more complex than the simple task. Although support for this suggestion may be drawn from the finding that all subjects tended to complete more of the complex symbols than the simple symbols, there are several reasons to doubt its validity. First, the finding that more complex than simple symbols were completed, irrespective of all other variables, probably results primarily from the fact that it takes more time to draw the simple symbols.[4] Second, the data on recall indicated that significantly ($p < .001$) more simple symbols were recalled than complex symbols, supporting the idea that the complex task did present subjects with a greater number of competing responses than the simple task. Additional evidence to support the notion that the complex task was in fact more difficult for subjects than the simple task may be inferred from the finding that recall of the complex task was disrupted by having another task succeed it, but recall of the simple task was unaffected by subsequent performance of a second task. Finally, a separate experiment was conducted at a later date to substantiate further the differential response competition of the two tasks. The tasks were presented as paired-associates learning tasks in which subjects wrote their responses to each number before the correct symbol appeared adjacent to that number in a memory drum. On the complex task subjects required significantly more trials to reach criterion ($t = 3.43$; $df = 17$; $p < .01$) and made significantly more errors ($t = 3.68$; $df = 17$; $p < .01$) than on the simple task.

The failure to find a Dissonance × Task Difficulty interaction on symbol completion suggests future modifications of the tasks employed as well as the time limits used. For instance, if the simple symbols were the roman numerals I, II, III, IV, V, and VI, the subjects might find them both easy to learn and easy to reproduce rapidly. Also, if some subjects were on 30-second time limits, some on 60-second time limits, and some on 90-second time limits or more, the hypothesized crossover effect derived from drive theory might be evaluated with greater sensitivity.

For the recall data also there was a failure to obtain the hypothesized Dissonance × Task Difficulty interaction. One reason for this might have been that on the simple task all subjects in the dissonant condition, regardless of order, were able to recall all six symbols, resulting in a mean recall score of 6.00. This suggests that the test for recall on the simple task was insensitive, providing a ceiling too low for adequate differentiation. It is not unreasonable to speculate that if the tasks consisted of more symbols a greater difference between the dissonant and consonant groups might have been obtained on the simple task than on the complex task, resulting in the predicted Dissonance × Task Difficulty interaction.

---

[4] To satisfy himself that this contention is valid, the interested reader may try to reproduce, as quickly as possible, the symbols used in the simple task, especially the symbols associated with 5 and 6. He will probably discover that these symbols take more time to draw than any of the symbols used in the complex task.

Although future research is obviously required to delineate further the dynamogenic effects of cognitive dissonance, the present investigation has demonstrated that the induction of cognitive dissonance enhances performance on a task which cannot be construed as dissonance reducing. In so demonstrating, this study has added validity to the conceptualization of dissonance as a construct possessing general drive properties.

REFERENCES

AMSEL, A. The combination of a primary appetitional need with primary and secondary emotionally derived needs. *Journal of Experimental Psychology,* 1950, **40,** 1–14.

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance.* New York: Wiley, 1962.

CARLSMITH, J. M., COLLINS, B. E., & HELMREICH, R. K. Studies in forced compliance: I. The effect of pressure for compliance on attitude change produced by face-to-face role playing and anonymous essay writing. *Journal of Personality and Social Psychology,* 1966, **4,** 1–13.

FARBER, I. E. Anxiety as a drive state. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1954.* Lincoln: University of Nebraska Press, 1954. Pp. 1–46.

FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *Journal of Experimental Psychology,* 1953, **45,** 120–125.

FESTINGER, L. *A theory of cognitive dissonance.* Stanford: Stanford University Press, 1957.

KELMAN, H. Attitude change as a function of response restriction. *Human Relations,* 1953, **6,** 185–214.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton-Mifflin, 1953.

MONTAGUE, E. K. The role of anxiety in serial rote learning. *Journal of Experimental Psychology,* 1953, **45,** 91–96.

RABBIE, J. M., BREHM, J. W., & COHEN, A. R. Verbalization and reactions to cognitive dissonance. *Journal of Personality,* 1959, **27,** 407–417.

RAMOND, C. K. Anxiety and task as determiners of verbal performance. *Journal of Experimental Psychology,* 1953, **46,** 120–124.

SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *American Psychologist,* 1958, **13,** 131–141.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning. *Journal of Experimental Psychology,* 1956, **52,** 296–305.

SPENCE, J. T., & SPENCE, K. W. The motivational components of manifest anxiety: Drive and drive stimuli. In C. D. Spielberger (Ed.), *Anxiety and behavior.* New York: Academic Press, 1966. Pp. 291–326.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychological Bulletin,* 1956, **53,** 303–320.

TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *Journal of Experimental Psychology,* 1952, **44,** 61–64.

ULLMAN, A. D. The experimental production and analysis of a "compulsive eating symptom" in rats. *Journal of Comparative and Physiological Psychology,* 1951, **44,** 575–581.

WEBB, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *Journal of Experimental Psychology,* 1949, **39,** 1–14.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation, 1955.

(Received August 29, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 132–138

ENERGIZING EFFECTS OF COGNITIVE DISSONANCE UPON DOMINANT AND SUBORDINATE RESPONSES[1]

NICKOLAS B. COTTRELL AND DENNIS L. WACK[2]
Kent State University

Is it appropriate to call cognitive dissonance a motivational variable? An experiment was conducted to determine whether cognitive dissonance energizes response tendencies in the manner of more traditional motivational variables. 15 university students committed themselves, without adequate justification, to perform a pseudorecognition task. 15 other students had adequate reason for performing the task. The task served to place previously established verbal habits in competition with each other. The results confirmed the hypothesis that cognitive dissonance enhances the emission of responses governed by strong habits at the expense of responses governed by weaker habits.

Festinger's (1957) theory of cognitive dissonance has inspired a great deal of significant research. Yet some psychologists feel that dissonance theory is isolated and disconnected from the rest of psychology. In their recent review of work on motivation, Cofer and Appley (1964) remarked, "Dissonance theory (and its research) needs more anchoring in the concepts of prior work than it has had to date [p. 798]."

Festinger defined dissonance as a motivational tension state with behavioral consequences identical to those of the homeostatic drives.

> In place of "dissonance" one can substitute other notions similar in nature, such as "hunger," "frustration," or "disequilibrium," and the hypothesis would still make perfectly good sense [Festinger, 1957, p. 3].

> Just as hunger is motivating, cognitive dissonance is motivating [Festinger, 1958, p. 70].

Many psychologists agree with Berlyne's (1964) statement,

> . . . to attack motivational problems means to seek factors that govern the organism's degree of alertness and activation, that bias the organism toward certain forms of behavior, and that determine what events will provide reinforcement for learning processes and how effectively [p. 448].

Although dissonance is defined as a motivational state, nearly all of the dissonance research has been concerned with demonstrating that dissonance biases the organism toward forms of behavior which reduce dissonance. There has been little work explicitly concerned with the activating effects of cognitive dissonance, and this deficiency is probably in large part responsible for the apparent disconnectedness of dissonance theory and research from the rest of psychology. Many psychologists feel that the energizing of responses is the most distinctive characteristic of motivational variables.

This feature of motivation receives particular emphasis in Spence-Hull theory (Brown, 1961; Spence, 1956, 1958). According to this view, motivational arousal energizes response tendencies regardless of whether these responses are instrumental to need satisfaction. Habit strength, rather than the instrumental value of the habit, determines which responses are enhanced by motivational arousal. The theory posits a multiplicative relationship between the drive and habit factors. When there is a hierarchy of competing responses, motivational arousal enhances the emission of responses governed by strong habits at the expense of responses governed by weaker habits.

To clarify the conceptual status of cognitive dissonance, the present study examined the assumption that the arousal of cognitive dissonance has energizing effects upon human performance similar to those ascribed to generalized drive (D) in Spence-Hull theory.

The present study used the insufficient justification paradigm and observed the effects of dissonance upon performance on a task shown by previous research (Zajonc & Nieuwenhuyse, 1964) to be sensitive to the energizing effects of drive. The present study tested the hypothesis that cognitive dissonance enhances the emission of responses governed by strong habits at the expense of responses governed by weaker habits.

---

[1] This research was supported by Grant GS-1016 from the National Science Foundation.

[2] The authors are grateful to Stuart Taylor for his thoughtful comments on an earlier draft of this paper.

While previous research (Brehm & Cohen, 1962) has shown that dissonance is aroused if a person voluntarily performs a task when there is little justification for doing so, only recently have the effects of insufficient justification upon task performance been investigated. Weick (1964) found that persons who voluntarily performed a task for reduced rewards evaluated the task highly, expended a great deal of effort, and performed successfully on the task. According to Weick, successful task performance aids in dissonance reduction by providing behavioral cognitive elements to support the enhanced evaluation of the task.

Subsequent studies of justification and productivity give some support to the assumption governing the present study by suggesting that changes in productivity are due in part to the energizing effects of dissonance. Ferdinand (1965) and others (see Weick, 1966, for a summary) found that voluntary participation for reduced rewards improved performance but did not increase liking for the task. These findings show that dissonance can act directly upon task performance without the mediation of attitudes toward the task. One interpretation of these findings is that performance improved because dissonance enhanced the emission of dominant task responses. This interpretation requires the assumption that the dominant task responses were the correct ones. Another interpretation, however, is possible. Perhaps, in some way that is not yet fully understood, successful task performance alone, without the mediation of task attitudes, is a means of reducing dissonance.

The assumption governing the present study is that cognitive dissonance energizes responses even when those responses are not instrumental to dissonance reduction. Since the findings of the studies discussed above are presumptive evidence that successful task performance alone might be a means of reducing dissonance, the present study employed a situation in which the subject could not control the proficiency with which he performed the task.

METHOD

Overview of Procedure

In the first stage of the experiment verbal habits of varying strengths were established. The dissonance manipulation was next. Then the verbal habits were placed in competition and the frequency of emission for the various responses was observed.

Task

Stimulus Materials

Seven-letter verbal stimuli were lettered on 4 × 6-inch cards in capital letters. The stimuli had a uniform length of 16.3 centimeters. The training stimuli were 10 of the three-syllable Turkish words used previously in recognition studies (Solomon & Postman, 1952; Zajonc & Nieuwenhuyse, 1964). Black and white 2 × 2-inch slides and 4 × 6-inch photos of the training stimuli were prepared. The stimuli for the pseudorecognition trials were nonsense words. Black and white 2 × 2-inch slides were made of the pseudorecognition stimuli.

Training

During training, verbal habits of varying strengths were established by manipulating the number of times different words were exposed. Five training frequencies were employed: 1, 2, 5, 10, 25. The 10 Turkish words were divided into five pairs. For each subject, one of the pairs was assigned to each of the training frequencies. Throughout the experiment the word pairs were rotated through the training frequencies in a 5 × 5 Latin square so that each pair occurred an equal number of times at each training frequency.

For each subject a deck of 86 photos was prepared consisting of 50 photos for the two 25-frequency words, 20 photos for the two 10-frequency words, 10 photos for the two 5-frequency words, 4 photos for the two 2-frequency words and 2 photos for the two 1-frequency words. Training was described to the subject as a study of how people pronounce foreign words. The photos were shuffled and exposed to the subject one by one. On each exposure the experimenter displayed the photo and read the word aloud. Then the subject pronounced the word once. The subject's pronunciation was neither reinforced nor corrected. The words were presented at a rate of one every 4 seconds, and training took about 5 minutes.

Pseudorecognition Task

This phase was described to the subject as an experiment on subliminal perception. The subject was told that the foreign words he had seen earlier would be presented one at a time and he was to call out the word presented. He was told that the speed of presentation would vary from trial to trial and that sometimes the word would be quite difficult to identify, and on those trials he was to make his best guess.

The stimuli were presented tachistoscopically on a screen 5 feet from the subject in four blocks of 53 trials each. If a subject failed to respond within 15 seconds, the experimenter urged him to make a response and did not show the next slide until he did so. It took about 40 minutes to complete the four trial blocks.

Each trial block consisted of 10 recognition trials and 43 pseudorecognition trials. The *recognition trials* of each block consisted of single presentations of each of the 10 training words. They were exposed with a medium diaphragm opening and a shutter speed of .2 second. Pretesting had shown that these settings produced 85% correct recognition. A random-number table was used to determine the location of the recognition trials within a block and the training frequency of the stimulus shown on a particular recognition trial. A different[2] schedule was prepared for each trial block. These schedules were held constant for all subjects, but since the training words were rotated through the training frequencies, the subjects differed with regard to the order in which the training words were presented on the recognition trials.

On the *pseudorecognition trials* nonsense words were presented at .01 second with a medium diaphragm opening. In pretests of these settings subjects reported seeing something wordlike, but they were unable to identify words and could not discriminate nonsense words from training words.

The subject's responses were recorded phonetically. The response was accepted if it matched a training word in two of its syllables. On the few occasions that the experimenter could not classify a response as one of the training words, the subject was told that this was not one of the foreign words he had learned and he was asked to make another guess. Subjects were excluded from the sample if their responses on the first block of trials indicated they had not learned at least two of the training words.

Dependent Variable

The dependent variable was the number of times words of each training frequency were emitted on the pseudorecognition trials. The pseudorecognition trials served to place the verbal habits established during training in competition with each other and to assess their relative strengths. Since the subjects were instructed to guess which training word was shown and since there was effectively no stimulus on the pseudorecognition trials, their guessing responses on these trials depended solely upon the habits established during training. It should also be noted that since there were no accuracy criteria on the pseudorecognition trials, the subject could not adjust the quality of his recognition performance on these trials.

Experimental Conditions

High-Dissonance Condition

The high-dissonance subjects participated in the experiment with little justification for doing so. Justification for participation was minimized by telling the subject that he could not get experimental points by participating in this experiment. At Kent State University, students in the introductory psychology course are required to win 6 experimental points. They are awarded 1 point for each ½ hour of experimental participation. The subjects in the present study had signed up for "Experiment III-C" which was listed as lasting 1½ hours and affording 3 points of experimental credit.

The procedure for denying points was similar to one used previously (Weick, 1964). The experimenter presented the subject a letter he had allegedly received from a committee of the psychology department. After the training phase the experimenter said to the subject:

> Before we start the second experiment there is something I suppose I should tell you. Last week I got this letter from Dr. H___ of the psych. department. I probably should have told you about it, but I've been so busy that I just didn't get around to it. Why don't I just read it to you so you'll know what it's all about.

The experimenter then read aloud the following letter:

> May 18, 1966
>
> Dr. Dennis L. Wack
> Department of Educational Research
> Kent State University
> Kent, Ohio
> Dear Dr. Wack:
>
> It has come to our attention that you have been using 162 students in Experiment III-C, and have been giving them points despite our letter to you of May 2 indicating that you would not be permitted to do so. I have conferred with Dr. B___[3] and Dr. G___, and they assured me that only members of the Psychology Department and their students are permitted to recruit subjects from 162 by offering points.
>
> Since you are affiliated with the Department of Educational Research, have had no experience in Psychology, and are doing research that from a psychological point of view is not particularly meaningful, you do not qualify to give experimental points. As of this date we are cancelling any points given for Experiment III-C. You will be permitted to set out sign-up folders in the Psychology Department, but you must clearly indicate to the students that they cannot earn points by participating in Experiment III-C.
>
> I trust that we have made our position clear. It must be recognized that 162 students are peo-

---

[3] These individuals were conducting the introductory psychology classes.

ple, not "guinea pigs" to be used indiscriminately for any study requiring human subjects. I can understand your wish to have subjects appear for your experiment. However, I would strongly advise that you continue it by obtaining subjects through the standard channels available to members of your department.

Sincerely,
J___ B. H___
Chairman, Committee on
Research
Department of Psychology

cc: Dr. B___
Dr. G___

After he had read the letter, the experimenter asked the subject to participate for no points with the following remarks:

> Here it is if you want to look at it. [Experimenter hands the letter to the subject.] Obviously I can't force you to stay. I suppose you could leave if you *have* to—undoubtedly there are other things you have to do at this hour—nevertheless, we could use your time. Besides it would be convenient for me if you would stay. How about it, shall we go ahead?

If the subject chose to stay, the experimenter administered the pseudorecognition instructions and task. The pseudorecognition task was not mentioned to the subject unless he chose to stay. The dissonance manipulation took about 7 minutes.

Low-Dissonance Condition

To control for recency of training and interference, the low-dissonance subjects were detained 7 minutes after training before receiving the instructions for the pseudorecognition task. The experimenter described the delay as a second experiment. He read aloud a passage from a history of the state of California. This passage was the same length as the letter read to the high-dissonance subjects. After he read the passage, the experimenter gave it to the subject and asked him to underline the topic sentence in each paragraph.

Questionnaire and End of Session

After the pseudorecognition task a questionnaire determined the subject's attitudes toward the task and toward his performance on it. The high-dissonance subjects were, in addition, interviewed to determine whether they had been suspicious of the manipulation and if they had felt free to leave upon learning they could not get points for participating. For the high-dissonance subjects the session concluded with debriefing in which they were informed of the purpose of the experiment, awarded points, and pledged to secrecy.

Subjects and Design

The subjects were 43 male introductory psychology students at Kent State University. There were 15 subjects in the high-dissonance condition. When interviewed, they all reported that they were aware they would not get points if they stayed and that they had stayed of their own volition. Twenty subjects were run in this condition and five were lost to the experiment. One subject failed to learn the training words, two chose to leave when they learned they would not get points, one was suspicious of the manipulation, and one was a member of a racial minority. There were 15 subjects in the low-dissonance condition. Twenty-three subjects were run and three were lost for failing to learn the training words. Time prevented running five more experimental subjects, so the last five low-dissonance subjects were discarded to equalize *N*s and simplify the statistical analysis.

The Latin square counterbalancing word pairs and training frequencies was completed three times in each experimental condition.

RESULTS

Pseudorecognition Task

Figure 1 presents the mean number of responses of each training-frequency class emitted in the two dissonance conditions on the pseudorecognition trials. There were 15 subjects in each condition and each subject contributed 172 responses (4 blocks of 43 pseudorecognition trials).

Figure 2 presents the mean number of responses of each training-frequency class emitted in the two dissonance conditions on each block of pseudorecognition trials. The data of Figure 2 were submitted to a three-factor (dissonance, training frequency, and trial blocks) analysis of variance. Three effects

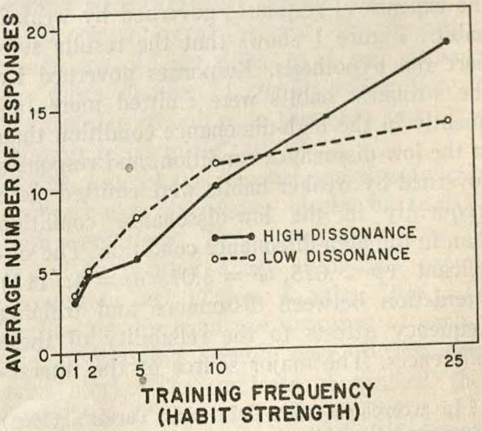


FIG. 1. Number of responses of different training-frequency classes emitted on the pseudorecognition trials, averaged over subjects and over trial blocks.

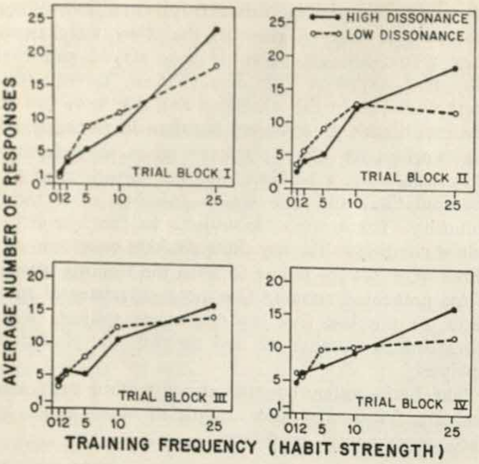


FIG. 2. Average number of responses of different training frequency classes emitted in separate blocks of pseudorecognition trials.

were significant. The main effect of training frequency is significant ($p < .001$, $F = 38.28$, $df = 4, 112$). Analysis of the trend components[4] reveals that the linear component of the training-frequency effect is significant ($p < .001$, $F = 99.46$, $df = 1, 28$), and accounts for 95% of the variation between training-frequency classes. Response emission on the pseudorecognition trials is an increasing linear function of training frequency. The slope of this function is .520.

The hypothesis of the present study predicts that dissonance enhances the emission of responses governed by strong habits at the expense of responses governed by weaker habits. Figure 1 shows that the results support the hypothesis. Responses governed by the strongest habits were emitted more frequently in the high-dissonance condition than in the low-dissonance condition, and responses governed by weaker habits were emitted more frequently in the low-dissonance condition than in the high-dissonance condition. The significant ($p < .025$, $F = 3.07$, $df = 4, 112$) interaction between dissonance and training frequency attests to the reliability of these differences. The major source of the interaction between dissonance and training frequency is the difference in linear trends between dissonance conditions. The linear component of the interaction accounted for 65% of the interaction sums of squares and is significant ($p < .05$, $F = 5.46$, $df = 1, 28$). The slope of the response emission function is greater in the high-dissonance condition (.642) than in the low-dissonance condition (.398).

Failure of the three-factor interaction of dissonance, training frequency, and trial blocks to achieve significance ($F < 1$) reveals that the effects of dissonance upon response emission were similar on all trial blocks. Figure 2 shows that on every trial block dissonance enhanced the emission of responses governed by strong habits at the expense of responses governed by weaker habits. Table 1 shows that on every trial block the slope of the response emission function in the high-dissonance condition was greater than the slope in the low-dissonance condition.

Table 1 also shows that on successive trial blocks the slope of the response-emission function for both dissonance conditions combined became less steep. This finding is reflected in the third significant effect of the analysis of variance, a significant interaction ($p < .001$, $F = 6.97$, $df = 12, 336$) between training frequency and trial blocks.

To summarize the results, response emission on the pseudorecognition trials is an increasing linear function of training frequency,

TABLE 1

SLOPE OF FUNCTION RELATING RESPONSE EMISSION TO TRAINING FREQUENCY FOR BOTH DISSONANCE CONDITIONS SEPARATELY AND COMBINED ON SEPARATE AND COMBINED TRIAL BLOCKS

Trials	Combined dissonance conditions	High dissonance (b_x')	Low dissonance (b_x)	$b_x' - b_x$
Block I	.774	.900	.648	.252
Block II	.481	.672	.289	.383
Block III	.479	.531	.427	.104
Block IV	.343	.463	.223	.240
Responses averaged over all blocks	.520	.642	.398	.244

Note.—There were 15 subjects in each condition. Response frequencies were averaged over subjects.

---

[4] In accordance with Gaito and Turner's (1963) recommendations, the analysis of trend components reported here used the models presented by Grant (1956). Gaito's (1965) suggestions were followed in deriving trend coefficients for unequal intervals.

and the slope of the response-emission function declines over successive trial blocks. As predicted, dissonance enhances the emission of dominant responses at the expense of subordinate responses.

Evaluation of the Task and Own Performance

The dissonance conditions did not produce differential liking for the pseudorecognition task (subliminal perception study). The conditions did not differ by *t* test in mean response to the following items: "Compared with other experiments, how interesting was the final study on subliminal perception?" and "Experiments like the subliminal perception study should be more frequent in psychology."

With one exception, the subjects in the two conditions did not differ in their perceptions and evaluations of their performance on the pseudorecognition task. The conditions did not differ by *t* test in mean response to five items: perception of own accuracy of recognition, perception of effort expended, report of how hard they tried to recognize the words, report of feeling bothered when failing to identify a word, and expected accuracy of recognition if they were to do the task again. The dissonance manipulation produced a difference which approaches the conventional significance level on one item: "During the subliminal perception study, how important was it to you that you identify the words correctly?" The high-dissonance subjects indicated that it was less important to them to identify the words correctly than did the low-dissonance subjects ($t = 1.714$, $df = 27$, $p < .10$).

DISCUSSION

The results of the present study support the view that cognitive dissonance has nonspecific energizing effects upon human performance. On the pseudorecognition trials cognitive dissonance enhanced the emission of dominant responses at the expense of subordinate responses. Since correct recognition was impossible on these trials, these findings were not a manifestation of dissonance reduction through increased productivity.

The energizing effects of dissonance were not accompanied by any differences in the subjects' attitudes toward the task and toward their performance. Thus the results of the present study, like those of Ferdinand (1964) and others cited above, did not show the evaluative task-enhancement process described by Weick (1964). These findings support the view that a discrepant commitment to perform a task need not produce task enhancement, and that task enhancement is not required for a discrepant commitment to affect a person's task behavior.

On one attitude measure, however, the difference between dissonance conditions approached the conventional significance level. The high-dissonance subjects attached less importance to correctly identifying the words on the pseudorecognition task than did the low-dissonance subjects. This nearly significant finding may indicate that efforts to reduce dissonance centered on psychologically revoking the commitment to participate, and would then permit an alternative interpretation of the response emission results. Rather than demonstrating the nonspecific energizing effects of dissonance, these findings may have resulted from the high-dissonance subjects' performing the task in a careless and perfunctory fashion in order to provide behavioral support for their view of the task as unimportant. There is little support for this interpretation, however. Careless and perfunctory responding on the pseudorecognition task would be characterized by inaccurate responding on the 40 trials on which the training stimuli were presented supraliminally and by rapid completion of the task. The dissonance conditions did not differ by *t* test either in the mean number of correct recognitions of the training stimuli or in the mean number of minutes to complete the task.

The slope of the response-emission function declined over successive blocks of trials in both conditions. The same result was obtained in another study using the pseudorecognition task (Zajonc & Sales, 1965). The decline results from the exposure of each of the 10 training stimuli once during every trial block. The subjects recognized these stimuli about 80% of the time (75.2% in the high-dissonance condition and 82.7% in the low-dissonance condition). The subjects were thereby reminded of all of the training words and could use them as guesses on the pseudorecognition trials. This procedure was especially helpful to responses assigned a low frequency in training, since they were most readily forgotten. As a consequence of seeing these low-frequency words, the number of times they were used as guesses increased. Since each subject had a constant number of trials, an increase in the use of low-frequency words as guesses would necessarily reduce the number of times the high-frequency words were used. Thus the slope of the response-emission function was lowered and its origin increased.

Numerous demonstrations that cognitive dissonance biases individuals toward behaviors which reduce or avoid dissonance do not permit the conclusion that cognitive dissonance is a motivational variable. If there is scientific utility in attempting to integrate dissonance theory and research with the rest of psychology, then research on the energizing effects of dissonance must continue. Another question awaits investigation. Is the reduction of cognitive dissonance reinforcing in that it forms and strengthens habits?

Attention to the energizing effects of dissonance may enable more specific predictions concerning which mode of dissonance reduction is employed in a particular situation. Potentially effective dissonance-reduction modes may be used seldom because the appropriate habits are weak relative to other habits activated by the situation and because cognitive dissonance decreases the probability of dissonance-reducing responses by enhancing the emission of the dominant responses.

REFERENCES

BERLYNE, D. E. A decade of motivation theory. *American Scientist*, 1964, 62, 448–453.

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance*. New York: Wiley, 1962.

BROWN, J. S. *The motivation of behavior*. New York: McGraw-Hill, 1961.

COFER, C. N., & APPLEY, M. H. *Motivation: Theory and research*. New York: Wiley, 1964.

FERDINAND, P. R. The effect of forced compliance on recognition. Paper read at Midwestern Psychological Association, Chicago, May 1965.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L. The motivating effect of cognitive dissonance. In G. Lindsey (Ed.), *Assessment of human motives*. New York: Rinehart, 1958. Pp. 65–86.

GAITO, J. Unequal intervals and unequal *n* in trend analysis. *Psychological Bulletin*, 1965, 63, 125–127.

GAITO, J., & TURNER, E. Error terms in trend analysis. *Psychological Bulletin*, 1963, 60, 464–474.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychological Bulletin*, 1956, 53, 141–154.

SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition threshold for words. *Journal of Experimental Psychology*, 1952, 43, 195–210.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale University Press, 1956.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *American Psychologist*, 1958, 13, 131–141.

WEICK, K. E. Reduction of cognitive dissonance through task enhancement and effort expenditure. *Journal of Abnormal and Social Psychology*, 1964, 68, 533–539.

WEICK, K. E. Task acceptance dilemmas: A site for research on cognition. In S. Feldman (Ed.), *Studies of cognitive interaction*. New York: Academic Press, 1966. Pp. 225–255.

ZAJONC, R. B., & NIEWENHUYSE, B. Relationship between word frequency and recognition: Perceptual process or response bias? *Journal of Experimental Psychology*, 1964, 67, 276–285.

ZAJONC, R. B., & SALES, S. M. Social facilitation of dominant and subordinate responses. *Journal of Experimental Social Psychology*, 1965, 2, 160–169.

(Received October 1, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 139–149

BARGAINING, EXPECTATIONS, AND THE PREFERENCE FOR EQUALITY OVER EQUITY [1]

WILLIAM R. MORGAN AND JACK SAWYER

University of Chicago

When, as frequently occurs in interaction between 2 persons, the possible rewards for one exceed those for the other, it might be expected that the outcomes they actually achieve would reflect this disparity. Surprisingly, however, between boys who are classmates this tended not to be true. Instead, strict equality was highly preferred—and at least as much so between those who were not friends as between those who were. An important part of resolving conflict lies in determining just what the other expects. Such information, as this experiment shows, facilitates, for both friends and nonfriends, the resolution of conflict—but in different ways. Friends, though they prefer equality, will accept inequality when one thinks the other might want it; for them, information, by reducing such well-intentioned misunderstanding, increases equal outcomes. Nonfriends appear less willing to accept anything but equality and reach it as surely with or without information; but knowing what the other expects lets them come to agreement in just $\frac{1}{5}$ of the time.

Any interaction where what one person most prefers differs from what the other most prefers presents a conflict of interest: both cannot get exactly what they want. Among the responses to such conflict are dominance, submission, and withdrawal; but another approach to resolving conflict—particularly promising for the long-term welfare of both parties—is mutual accommodation, a process that is represented by explicit or implicit bargaining.

In bargaining, the object is to agree upon actions that produce a mutually acceptable outcome. In seeking an acceptable outcome, it is easy to discard from consideration outcomes that give either party less than he could get by himself, acting independently, as well as outcomes that neither person favors. But this cooperative process leads inevitably to a set of outcomes where interests are exactly opposed: the better an outcome is for you, the worse it is for the other. It is among this set of outcomes that bargaining occurs—and usually there is no single outcome that appeals to both parties as the obvious solution. In the general bargaining situation there are several outcomes, rarely symmetrical between the two persons, and outcomes are reached by joint agreement rather than by independent choice.

Reaching agreement means identifying a mutually acceptable outcome. What is acceptable, however, depends upon the expectations each person has. But at the onset of interaction, your estimate of what the other expects is usually imprecise, and so expectations usually change over the course of interaction. Starting from these imperfectly known expectations, the process of reaching agreement can be characterized as essentially a mutual adjusting of expectations (Harsanyi, 1962; Iklé & Leites, 1962); agreement comes when your expectations and those of the other are congruent.

Often this adjusting process involves many offers and counteroffers, as each person tries to learn, and possibly modify, the other's expectations. For example, in bargaining over how much to pay for a used car (where, as in social interaction more generally, intrinsic values are often difficult to assess), what you most need to know is just how little the seller will take (and likewise, what he needs to know is just how much you will give). Thus, a key to successful bargaining in such situations, as Kelley (1966) suggests, lies in finding out the other's expectations while concealing your own. For if you know the least he will take, you can represent this as exactly the most you will give, and thus obtain the most favorable outcome possible. So you need to know the other's expectations but are reluctant to tell him yours, and he is similarly disposed; in these conflicting needs, as Kelley indicates, lies a central problem of bargaining.

---

[1] This article is based upon research conducted by the senior author for his AB honors thesis at the University of Chicago, June 1965. Support was provided by United States Public Health Service Grant MH 05350, "Experiments in the Resolution of Interpersonal Conflict." The authors are grateful to Peter M. Blau for his insightful comments on the findings.

The above analysis assumes, as nearly all analyses of bargaining have, that you are interested strictly in maximizing your direct return on this particular occasion. Wise persons, however, do not treat every encounter like a used-car transaction where they never expect to see the other again; friends, instead, also take care that the other obtains an outcome sufficiently rewarding so that he is willing to interact again. Here, too, however, knowledge of what the other expects is crucial—perhaps even more so, as the comic and tragic misunderstandings of the well intentioned suggest. Thus both in formal and friendly relations, expectations should be important, but the way they operate may differ—as this research will show—for those who are friends and those who are not.

The general thesis of this research is that information about what the other expects facilitates bargaining, and this is tested in hypotheses that examine the effect of such information in the two major aspects of bargaining—the eventual outcome and the process by which it is reached. Such information should enable bargainers to reach an outcome that is more mutually satisfactory, and to reach it more quickly.

The comprehensive analysis of process—though the heart of explanation—is complicated indeed, and includes the interaction of initial expectations, initial and return bids, the succeeding sequence of bids, changing expectations, and final outcome. In a sense the most basic and summary measure of process in the resolution of conflict, however, is simply how much time bargaining consumes. If obtaining information about the expectations of the other person forms a major problem in the bargaining process that must be solved before an outcome is reached, then one would expect to find

Hypothesis 1: Knowing what the other expects will shorten the duration of bargaining.

Initial ignorance of the other's expectations not only calls for time to explore them, but also means they may never be fully known. Thus, complete knowledge of expectations should change the outcome as well as the duration. In general, more information should produce an outcome that more fully reflects the strengths and desires of the two persons, and hence one that is more stable.

Since the outcomes on which bargaining eventually focuses may be ordered so that the more one person gets, the less the other gets, there are just two directions in which information can move the outcome. The two directions may be identified by two prominent outcomes among the ordered set: toward *equality*, where the two get absolutely the same amount, and toward *equity*, where the two get proportionally the same, each relative to the possibilities he has (so that the person with the stronger position gets more). (In the special—but most frequently investigated—case of exact symmetry between two persons, equity and equality coincide; the present research, by employing an asymmetric situation, permits distinguishing these rather different outcomes.)

Game theory (Nash, 1950, 1953) specifies the "equitable" outcome, which Harsanyi (1961) reasoned rational bargainers having full information should always reach. Thus, when the strengths and desires of the two persons tend to make them rational bargainers in the game-theory sense, information should promote the equitable outcome. Such a condition is represented in this experiment by two persons who are *not* friends (presumably leaving each, like the buyer and seller in the used-car transaction, free to pursue his own immediate interest), with the stronger position of one legitimated by higher status.

The opposite condition is represented by two persons who *are* friends, with the stronger bargaining position of one *not* legitimated by higher status. Such parties conform less to the model of rational bargainers seeking only their immediate self-interest; here the effect of information should be to move the outcome toward that more natural for friends—strict equality. Thus, whether knowing what the other expects moves the outcome toward equality or toward equity should depend on the relation that exists between the two persons: hence,

Hypothesis 2a: Knowing what the other expects will promote equal outcomes between friends, and

Hypothesis 2b: Knowing what the other expects will promote equitable outcomes between nonfriends.

METHOD

Subjects

The 56 subjects were fifth- and sixth-grade boys between the ages of 10 and 12. Subjects of this age were chosen to help assure that they would be strongly motivated by the reward itself. For a boy who receives a weekly allowance of 50¢ or less (as two-thirds of these do), the chance to win several weeks' allowance in a few minutes surely means more than does a chance at the same $1.50 to a college student. (Girls were not used because pretrials suggested that they were less strongly and uniformly motivated by the monetary reward itself.)

The Bargaining Situation

Two boys faced each other across a Bargaining Board (Figure 1), where each had in front of him seven pockets, representing seven alternative outcomes. For one boy, the pockets contained nickels, increasing successively from none at one end to six at the other end. For the other boy, the pockets contained from zero to six quarters—but in exactly the reverse order. The two boys sought to agree, through offers, counteroffers, and discussion, on one of the seven pairs of directly opposite pockets. When they agreed on a pair of pockets, each got the money from the pocket on his side; if they failed to reach agreement, neither would get anything. The two boys were not allowed to split the money afterwards. The experimental procedure prevented boys from knowing whom they would play until their turn to participate, and otherwise attempted to minimize collusion—implicit or explicit. Instructions further discouraged any hope by the boy playing for nickels that if he let the other get a lot, the money could be shared. The success of these measures is suggested by the absence, over the course of the experiment, of any increase in unequal outcomes.

The Bargaining Board represents the most basic characteristic of a bargaining situation—a set of alternative outcomes in which the more one party gets, the less the other party gets. Three features of this situation particularly distinguish it from most conflict experiments: First, parties to a conflict reach resolution through communication leading to joint agreement, rather than through independent choice; second, rather than having only two choices, subjects have a range of alternatives; third, the relation between the two persons is not symmetrical, so that the alternatives and outcomes one faces are not exactly the same as those the other faces. Such features—joint bargaining, multiple alternatives, and asymmetry—are entirely usual in real-life conflict situations.

Asymmetry has the special characteristic of making the resolution of conflict nontrivial. Otherwise—if, for example, both boys had quarters, so that the middle pocket gave each 75¢—there would be one overwhelmingly attractive outcome. When a single outcome is highly prominent, the possibility of conflict is greatly reduced (Schelling, 1960). Asymmetry, however, permits distinguishing three prominent outcomes: *equality*, giving each person 25¢; *equity* (the game-theory solution), giving each half his maximum—75¢ to one and 15¢ to the other; *maximum sum* (the "social welfare" solution for this two-person society), giving $1.50 to one and nothing to the other.

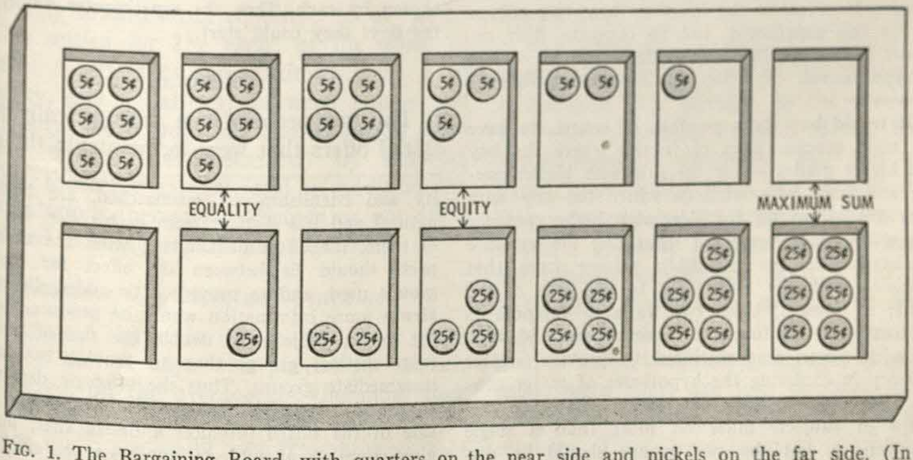


FIG. 1. The Bargaining Board, with quarters on the near side and nickels on the far side. (In practice, the words in this figure do not appear on the board—only the coins themselves.)

Pairing Friends and Nonfriends

Friendship was assessed by asking each boy to name the five other boys in his classroom whom he would most like to have sit next to him (presuming that a new seating arrangement were to be made) and the five he would least like to have sit next to him. Preferences from three classrooms were used to constitute 28 pairs, each consisting of two boys from the same classroom. Fourteen pairs were composed of boys who had listed each other among those they would *most* like to sit next to, and the other 14 pairs were composed of boys who had listed each other among those they would *least* like to sit next to. The former pairs are called "friends," the latter, simply "nonfriends," since they need not be enemies, and certainly not strangers. It seems highly likely that by the spring, when this experiment was conducted, each boy knew every other boy in his classroom.

Differences in recently received grade reports were used to accentuate the distinction, in the two sets of pairs, between the more friendly relation which was expected to lead toward equality, and the more formal relation which was expected to lead toward equity. Each pair was constituted so that one boy had considerably higher school grades than the other; the difference in grade rankings between the two boys was equal, as nearly as possible, to half the number of boys in their class. Then boys were located on one side of the board or the other so that for nonfriends, but not for friends, the difference in grades was consistent with the difference in payoffs. For nonfriends, the boy with higher grades always sat on the side with higher payoffs; for friends, he always sat on the side with lower payoffs.

This made the a priori relation between friends more nearly equal and that between nonfriends more unequal, as these relations in practice tend to be. Thus this manipulation supplements, in a consistent way, pairing by seat-mate preferences, and helps compensate for possible weaknesses in that single measure. It promotes the function these two groups play in this experiment: not to compare their responses themselves, but to provide groups for which, as hypothesized, the effect of information differs.[2]

To the boys, assignment to a side of the board (placed between them after they were seated) was apparently indiscriminate; it was not the purpose, among nonfriends, to *tell* the boy on the high side he was there because of better grades, but rather to give him a way, *if he chose,* to legitimize his better position. To this end, the experimenter, before bargaining started, read aloud the answers of both boys to a few background questions they had just completed; after reading both responses to a question about their most recent grade report, he casually commented that boys with high grades usually do well in this game. Thus for nonfriends, but not for friends, it was possible to legitimize the differences in payoffs by differences in grades. As it turned out, however, grades were almost never mentioned, either in bargaining, or in postinterviews.

Information Versus No Information

The experimental treatment, administered to a random half of the pairs in each of the two groups, was the giving or withholding of information about what the other expected. Expectations were solicited, in all pairs, by a questionnaire administered (in such a way that neither boy could see the other's answers) just after instructions were given and before bargaining started. On it, each boy answered three questions: "What do you think is the *most* you could win, if you tried hard?" "What do you think is the most *he* could win, if he tried hard?" "What do you think is the *lowest* amount you might win?" The crucial questions were the first and third (the second was asked mainly to facilitate answering the third). These two questions provided upper and lower limits for each boy's expectations and established for each an interval similar to what Iklé and Leites (1962) called the bargaining range. In the information condition, the experimenter read aloud the answers to these two questions; in addition, so this information would appear before the boys throughout their bargaining, he placed bold-lettered "most" and "least" markers on the pockets so chosen by each. Then the experimenter simply told the boys they could start.

RESULTS

The two persons in a pair typically made initial offers that were inconsistent; the aver-

---

[2] It would have been possible, of course, to have two more groups: pairs of friends where the boy with higher grades sat on the side with higher payoffs, and pairs of nonfriends where the boy with lower grades sat on the side with higher payoffs. This would have permitted separating the variance associated with the friendship pairing from that associated with the assignment to one side of the board; at present, these two variance components are completely confounded. Their separate identification, however, was sacrificed to provide greater efficiency in exploring the hypotheses of major concern. For this, it does not matter that the two groups of subjects differ on more than a single characteristic, but the two groups should be distinct. The two other potential groups, where equality and friendship are *mis*matched, are both less natural and less distinct from each other. For both of them, the effect of knowing what the other expects should lie between the effect for the two groups used, and so provide little added distinction. Hence more information would be produced by using added subjects to double the size of the two more distinct groups than to provide two added, intermediate groups. Thus the strategic decision to confound friendship pairing and assignment to a side of the board produces a design that, for any given number of subjects, appears more efficient in investigating the hypotheses in question.

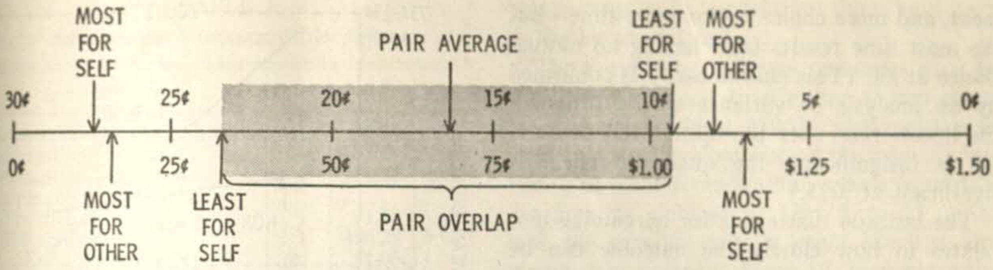


FIG. 2. Average expectations of 28 boys playing for nickels (shown above the payoff scale) and of 28 boys playing for quarters (shown below the payoff scale).

age disparity was 3.3 pockets on the Bargaining Board, where the maximum disparity is 6. Then they alternatively made successive offers, as quickly or as slowly as they liked, until they reached agreement. (One pair, after 30 minutes, was told their time was up, and immediately came to agreement.)

In the process of bargaining, subjects usually made brief comments to one another, such as "This is the fair one, we both get the same," or "Take this one, it's better than nothing." The number of offers required for agreement averaged 17, but ranged from 2 to 46. The duration of bargaining averaged 9 minutes, and ranged from 1 minute to 30. Duration correlated .82 with the number of offers, and appears to be a somewhat more standard and central measure to analyze; it probably reflects better the difficulty of reaching agreement, since the rate of offers seems to be in part a function of personality style.

The outcomes eventually reached tended strongly toward strict equality: 18 of the 28 pairs settled for 25¢ each; of the other 10 pairs, 3 settled at the maximum sum, 1 at equity, and 6 half-way between equity and equality. Both outcome and duration were affected (as later described) by knowing what the other expects, and in addition the expectations themselves—known or not—also related to both outcome and duration.

Expectations, Duration, and Outcome

Figure 2 shows, in its top half, how much money the 28 boys who played for nickels expected, on the average, as the most they could get, the least they could get, and the most the other could get. The bottom half shows the average expectations of the 28 boys who played for quarters. The expectations appear generally plausible, and the average estimate of the most the other might get agrees closely with the average the other actually reported as the most he himself expected (though in individual pairs, of course, larger disparities occurred). There appears to have been an optimistic bias, however, between what the boy playing for quarters estimated was the least he might get (33¢), and the amount (16¢) implied by his estimate of the most the other could get.

The major significance of these various expectations is in their overlap, which is considerable, and in their means, which lie toward the end that favors the person on the low side. The most crucial expectation is the least each person will take, and the difference and average of these values for the two persons in a pair produce the measure of "pair overlap" and "pair average." Means for both of these measures, over all 28 pairs, are also shown in Figure 2.

Pair overlap is important because it determines the initial latitude two persons have to bargain; this latitude, as the shaded area of Figure 2 shows, averages almost half the entire range of outcomes. But this overlap varies considerably from pair to pair, and relates strongly to the length of time a pair requires to come to agreement. When there are initially *no* pockets both would accept, bargaining requires an average of 15 minutes; when there is a single pocket, or just two, or three, bargaining averages only 4 minutes; when there are four or more mutually acceptable pockets, bargaining averages 9 minutes. Thus, bargaining is quickest when there is room, but barely room, for agreement, and more choice means more time—but the most time results from having no mutual choice at all. (This curvilinearity is confirmed by an analysis of variance where, although the linear trend over the three levels of overlap is insignificant, the quadratic effect is significant at .025.)

The latitude that exists for bargaining also relates to how closely the outcome can be predicted from the expectations. In general, higher expectations lead to higher outcomes; the correlation between pair average and the outcome is .53 ($p < .01$). But the accuracy of prediction varies with the amount of overlap. When there are no more than three mutually agreeable pockets, most outcomes (seven out of nine) lie *within* one pocket of the average expectation; but when there are four or more mutually agreeable pockets, most outcomes (13 out of 19) lie *more* than one pocket from the average. Thus the narrower range of acceptable outcomes makes the outcome more predictable ($p = .028$ by Fisher exact test, one-tailed).

Altogether, then, expectations predict both duration and outcome of bargaining; thus, they help illuminate the experimental results that follow, by suggesting the process through which the causal effects operate.

Effect of Information on Outcome

As predicted (Hypothesis 2a), knowing what the other expects, when the other is a friend, substantially increased the number of equal outcomes. Without information, the mean outcome (Figure 3) approximated the game-theoretic solution; with information, it closely approached strict equality. The significance of this result was examined by analysis of variance, not of the overall effect of information, but of its simple effect (Winer, 1962, pp. 232–238) separately for friends and nonfriends, since particular effects were separately hypothesized for the two groups. (Note that Hypotheses 2a and 2b are not merely what the Groups × Information interaction would test—that the effect of information differs for the two groups.) Rather, they specifically hypothesize the direction of the effect for each group. The result of the analysis of variance shows the effect of information for friends to be significant at the .05 level. Between nonfriends, however, information did not have the predicted effect (Hypothesis 2b); instead—with or without information—strict equality proved unexpectedly attractive.

TABLE 1

ANALYSIS OF VARIANCE OF OUTCOME

Source of variation	*df*	*MS*	*F*
Groups: Friends vs. nonfriends	1	1.29	.56
Simple effect of information			
For friends	1	10.29	4.50*
For nonfriends	1	.29	.13
Error	24	2.29	

*$p < .05$.

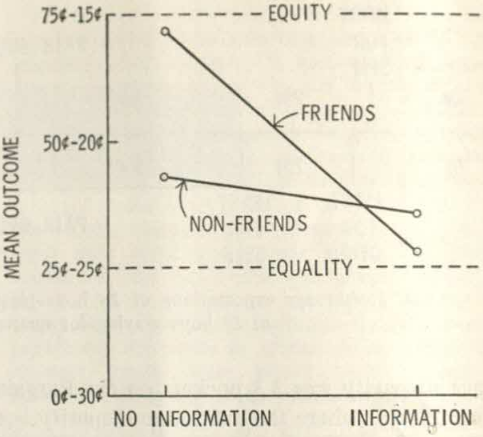


FIG. 3. The effect of information on outcome.

Effect of Information on Duration

The effect of information differs, as the significant interaction in Table 2 shows, for friends and for nonfriends. This was unexpected—unlike the different effects of information on outcome, which were directly hypothesized. Nonetheless, the significant interaction leads to the same procedure: examination of simple effects of information separately for the two groups. These effects, also presented in Table 2, simply represent a different partitioning of the variance ascribed to Information and Information × Groups. Note that the mean squares (which equal the sums of squares, since $df = 1$) add to 511.78 for both pairs of effects.[3]

---

[3] As a methodological check, these two analyses of variance were both also computed with the dependent variable transformed by taking its square root.

TABLE 2

ANALYSIS OF VARIANCE OF DURATION

Source of variation	df	MS	F
Groups: Friends vs. nonfriends	1	1.75	.02
Information	1	132.89	1.64
Information × Groups	1	378.89	4.67*
Simple effect of information			
For friends	1	31.50	.39
For nonfriends	1	480.28	5.93**
Error	24	81.05	

*$p < .05$.
**$p < .025$.

The results show that, for friends, where information was found to affect outcome, it did not affect duration; but for nonfriends, knowing what the other expected drastically reduced the time required for agreement. Without information the bargaining—generally lively and often heated—required an average of 15 minutes; with information, agreement came in an average of 3 minutes (Figure 4).

DISCUSSION

The Attractiveness of Equality

Among boys of this age—friends or not—strict equality, it appears, is a highly attractive outcome. Its attraction is strong enough that a boy with much better possibilities is willing to forego them for the sake of equality, and is expected to do so by the other. Effectively, these boys tended to define their situation as much more nearly symmetrical than the formal structure the experimenter imposed. Factors promoting symmetry overwhelmed the experimental asymmetry that gave one boy the chance to win five times what the other could. Notably, symmetry proved stronger than asymmetry even when the more favored boy had higher grades, and the two were not friends. If a symmetrical outcome is attractive to both friends and nonfriends, the reason may lie in the relation these boys have in the larger setting.

The prevalence of equal outcomes may be considered to result from either of two orientations, which Piaget (1965) distinguishes as stages in the development of the idea of justice. The earlier orientation, which is said to characterize children of ages 8 to 11, is to look for equality only in exact identity. Beginning at about age 11 or 12, however, "the child no longer thinks of the equal rights of individuals except in relation to the particular situation of each [p. 317]"—that is, considerations of equity obtain. An orientation of equity, though, can still produce an equal outcome, if—as appears to have occurred in the present study—the "particular situation" of each is considered equivalent.

Within the range of ages (10–12) in this research, preference for equality is uniformly strong. This indicates that even if, as Piaget found, considerations of equity do obtain, they also lead to an equal outcome. In turn, this implies that the asymmetry of the experimental situation is less important than the symmetry of the larger situation. Even considering their unequal experimental positions, boys generally regarded equality as the fair outcome. Interpreted according to Homans (1961), this means that *distributive justice* calls for equal *profits* because the relevant *investments* of each (age, status, etc.) are perceived (despite the experimental groupings) to be equal. In Adams' (1963) termi-

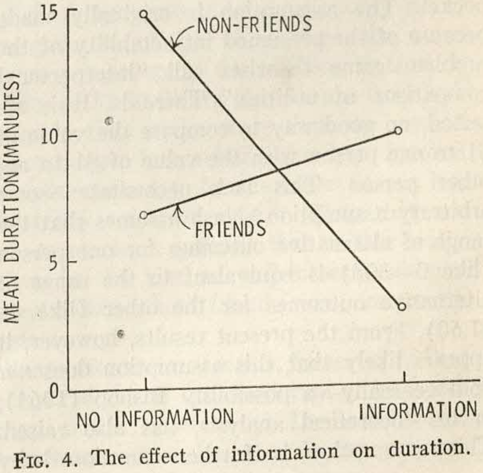


FIG. 4. The effect of information on duration.

This not only promoted homogeneity of error variance and normality of within-cell distributions, but also provided scales that probably correspond more closely to psychological magnitudes: the difference between durations of 1 minute and 5 should appear larger than the difference between 21 minutes and 25. The F values were generally somewhat higher than with the raw data, but the differences were not great, and so the more direct, untransformed analyses are presented.

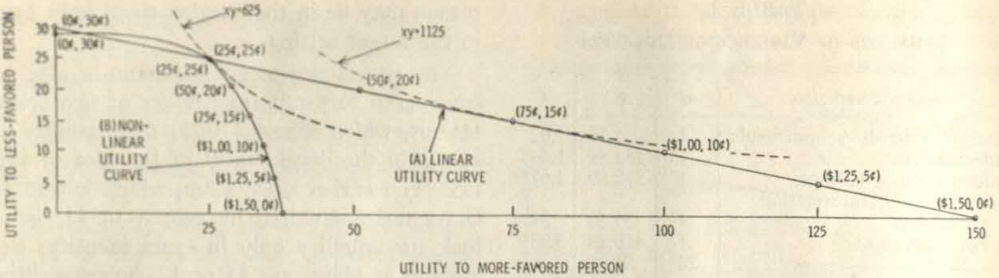


FIG. 5. Utility to two persons of various monetary outcomes: (*a*) where utility is linear with money, and (*b*) where it is not.

nology, equity calls for equal *outcomes* because the *inputs* are perceived to be equal.

The assumptions of the game-theory solution. The preponderance of strict equality has implications for the game-theoretic rationale (Nash, 1950, 1953) that specifies an outcome (equity) proportional to the opportunities each possesses. Assuming that the utilities these amounts of money have for the subjects are not *too* disproportionate to their actual monetary values (a possibility explored later), then at least one of Nash's four assumptions must not hold. The assumption most vulnerable here appears to be that of transformation invariance—that the solution will not be changed when an order-preserving linear transformation is made of either person's utilities.

This assumption implies that if the middle pocket is *the* solution when *both* persons have nickels (or both have quarters), then multiplying one person's outcomes by 5 (or 500) should leave the solution still at the middle pocket. The assumption is originally made because of the presumed intractability of the problem game theorists call "interpersonal comparison of utilities." There is, it is asserted, no good way to compare the value of $1 to one person with the value of $1 to another person. This lack necessitates some arbitrary assumption; Nash assumes that the range of alternative outcomes for one person (like 0—30¢) is equivalent to the range of alternative outcomes for the other (like 0–$1.50). From the present results, however, it appears likely that this assumption does *not* hold generally—a possibility Bishop (1963), in his theoretical analysis, has also raised. This assumption is further questioned by findings of another experiment[4] where bargainers who know the values of the other person's alternative rewards reach different outcomes (farther from the Nash solution) than those who do not.

Instead of accepting an arbitrary assumption about utilities, one can try to measure them. Measures suggesting the value of money include a boy's allowance, or a man's income, or wealth, or how long he would work for $10. Preferences in experimental gambles (Davidson, Suppes, & Siegel, 1957) have been used to assess utility. Though clearly fallible, such measures may well reflect comparative utilities more closely than the present game-theoretic assumption.

The utility of money. It is possible that the failure of the Nash solution to predict the attractiveness of strict equality lies not in the weakness of one of his four assumptions, but in differences between the face value of money and the utility persons place upon it. As Bernoulli (1738) pointed out, people generally have a declining marginal utility for money: the second million is hardly ever worth quite as much as the first, and, for these boys, the second 25¢ may mean less than the first 25¢. This makes the intervals between pockets unequal and may cause the maximum product of the two utilities, which identified Nash's solution, to fall on some pocket other than the middle one.

Figure 5 illustrates how the Nash solution differs according to the relation between money and the utility placed upon it. The straight line depicts a relation where the utility of money is strictly proportional to its actual value, and should approximate the shape of the utility curve, for these small amounts of money, among those of such relative affluence as the readers of this journal. (The interval scales of utility (not money) comprising the ordinate and abscissa are arbitrary in origin and unit, and have been fixed so that, for both persons, 0¢ has zero utility, and 25¢ has 25 units utility—the same for each.)

---

[4] W. E. Groves & J. Sawyer. How information helps the underdog. Social Psychology Laboratory, University of Chicago, April 1967.

For the linear utility curve, the hyperbola $xy = 1{,}125$, by being tangent at the (75¢, 15¢) outcome, reflects that for no other outcome is the product of utilities as high. That same point would still have the maximum product if any transformation of the form $x' = ax + b$, where $0 < a < \infty$, were made of either or both person's utilities.

It is possible, however, to construct an alternative, nonlinear utility curve where the maximum product occurs for the (25¢, 25¢) outcome—as the hyperbola $xy = 625$ tangent at that point of the alternative curve shows. Still, as can be noted, to construct a utility curve where no other outcome (particularly, 50¢, 20¢) produced as high a product as (25¢, 25¢) required a very major departure from linearity—with utilities such that $1.50 is worth only half again more than 25¢.

Yet many lesser departures from linearity—though themselves substantial—would *not* move the Nash solution from (75¢, 15¢). Consider, for example, a power function, $x' = x^m$, such as Stevens (1962) has generally found, across a wide range of sense modalities, to relate sensation to physical magnitude. This transformation would leave the Nash solution unchanged, since for any power—square root, cube, tenth root, whatever—if $x_1y_1 > x_iy_i$, where i is any other outcome, then $x_1{}^my_1{}^m > x_i{}^my_i{}^m$, where $m > 0$. Thus it is not at all clear that relatively simple, order-preserving transformations of a person's money values will alone account for the prominence of the equal outcome. The concept of utility is certainly relevant, but the transformation from money to utility may have to be calculated so that the utility a person has for a certain outcome is a function not only of his reward, but of the reward of the other as well—as in the consideration of vicarious utility (Valavanis, 1958; Preston, 1961) or altruism (Sawyer, 1966).

Bargaining Between Friends

Of the four combinations of friendship and information, the only one in which there was a strong departure from the equal outcome was, as Figure 3 shows, one involving friends, rather than nonfriends. This unexpected result might at first be thought to represent the altruism that friends tend to show one another (Sawyer, 1966): the less-favored person could actually prefer to get a little less so that his friend would get a lot more.

Such a preference, however, should hold whether the other's expectations were known or not: an altruistic orientation prefers more for one's friend even (or perhaps particularly) if he does not insist upon it. Yet, when both persons knew the other's expectations, they did not agree to let the person with the better possibilities profit from them, but instead strongly preferred a strict equality. Further, friends who lacked information and failed to reach equality generally expressed, in interviews afterward, uneasiness with the outcome.

This preference for equality, when expectations are known, suggests that in the absence of information friends develop and act upon false expectations about the other's preferences. Given the structure of the rewards, it is usually rather obvious that the person on the low side would prefer equality—which three-quarters of the time is his opening offer. It is *not* obvious, however, what the person on the high side wants, but the other may think he wants a lot, particularly since, three-quarters of the time when the more favored person opens the bidding, it is with an offer of $1.50. When the two are friends—a relation that means more than the reward from a single interaction—the less favored person is inclined to let the other *have* a lot, if he really seems to want it. Thus they are likely to reach an outcome that neither really prefers, in the kind of well-intentioned misunderstanding that friends are frequently prone to.

What information reveals, however, is that the favored friend does *not* want to take advantage of his opportunity. The less favored person learns that the other in fact expects less than he might be thought to. The overlap in the least amount expected is 50% greater for friends than for nonfriends—and this greater overlap permits the outcome, in the absence of information, to depart as it does from equality.

Knowledge of what the other expects may of course be learned not only from formal statements at the onset of interaction, but also in its normal course over time. When formally presented information was lacking, the four of seven pairs of friends who *didn't* reach equality bargained an average of only 2 minutes. The three pairs who *did* reach equality, however, took an average of 13 minutes, presumably finding out what the experimenter didn't tell them—that they preferred equality.

Bargaining Between Nonfriends

For nonfriends, equality appears to be more than merely the preference it is for friends, but closer to a requirement. Unlike friends, nonfriends tend to reach equality even when they don't know what the other expects. What the lack of information means, though, is that much more time is required to reach agreement. Lacking specific information, it takes time to find out that only near equality is acceptable.

When each knows what the other expects, however, the smaller overlap of expectations that nonfriends have is revealed, and need not be learned through actions. Disclosure thus substitutes for the knowledge that otherwise would come through successive offer and counteroffer. If reaching agreement requires learning what the other will accept, then having this information should make bargaining easier, and so reduce the time. This should be true when—as seems more likely for nonfriends—interaction is primarily instrumental, directed mainly toward the money reward. Between friends, however, where interaction itself is rewarding (Homans, 1950), the results show no pressure to end the interaction as soon as the state of information permits.

In fact, as Blau (1964) suggests, too quick or too strict concern with reaching equal balance in an interpersonal exchange appears to characterize nonfriends more than friends. Friends should be willing to tolerate temporary imbalances, while nonfriends want to keep their relation on more of a businesslike basis. With nonfriends, relations are more formal, like a price system, and direct appeal to the norm of equality is more frequently made: the less favored person frequently argued that equality was the only fair outcome. Similarly, Thibaut and Faucheux (1965) indicated that the less powerful person should appeal to a norm of "fair sharing," where such a norm serves, as earlier theorized (Thibaut & Kelley, 1959) as a substitute for the personal influence the weaker party lacks.

The difference in bargaining between friends and nonfriends is again illustrated by the norm of equality. Though the norm operates for both groups, it must be invoked, with nonfriends, by the less favored person; with friends, however, the favored person actually appears to prefer equality more than the less favored person thinks he might.

Everything considered, in bargaining between boys who are classmates—friends or not—strict equality is highly attractive, and reaching it is facilitated by knowing what the other expects—but in different ways for the two groups: information lets *friends* approach more *closely* the equality they *prefer;* it lets *nonfriends* approach more *rapidly* the equality they *require.*

REFERENCES

Adams, J. S. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology,* 1963, **67**, 422–436.

Bernoulli, D. Specimen theoriae novae de mensura sortis. *Comentarii Academiae Scientiarum Imperiales Petropolitanae,* 1738, **5**, 179–192. (Trans. by L. Sommer, *Econometrica,* 1954, **22**, 23–36)

Bishop, R. L. Game-theoretic analyses of bargaining. *Quarterly Journal of Economics,* 1963, **77**, 559–602.

Blau, P. M. *Exchange and power in social life.* New York: Wiley, 1964.

Davidson, D., Suppes, P., & Siegel, S. *Decision making: An experimental approach.* Stanford: Stanford University Press, 1957.

Harsanyi, J. On the rationality postulates underlying the theory of cooperative games. *Journal of Conflict Resolution,* 1961, **5**, 167–178.

Harsanyi, J. Bargaining in ignorance of the opponent's utility function. *Journal of Conflict Resolution,* 1962, **6**, 29–38.

Homans, G. C. *The human group.* New York: Harcourt, Brace, 1950.

Homans, G. C. *Social behavior: Its elementary forms.* New York: Harcourt, Brace & World, 1961.

IKLÉ, F. C., & LEITES, N. Political negotiation as a process of modifying utilities. *Journal of Conflict Resolution,* 1962, 6, 19–28.

KELLEY, H. H. A classroom study of the dilemmas in interpersonal negotiations. In K. Archibald (Ed.), *Strategic interaction and conflict.* Berkeley: University of California, Institute of International Studies, 1966. Pp. 49–73.

NASH, J. F. The bargaining problem. *Econometrica,* 1950, 18, 155–162.

NASH, J. F. Two-person cooperative games. *Econometrica,* 1953, 21, 128–140.

PIAGET, J. *The moral judgment of the child.* New York: Free Press, 1965.

PRESTON, L. E. Utility interactions in a two-person world. *Journal of Conflict Resolution,* 1961, 5, 354–365.

SAWYER, J. The Altruism Scale: A measure of cooperative, individualistic, and competitive interpersonal orientation. *American Journal of Sociology,* 1966, 71, 407–416.

SCHELLING, T. C. *The strategy of conflict.* Cambridge, Mass.: Harvard University Press, 1960.

STEVENS, S. S. The surprising simplicity of sensory metrics. *American Psychologist,* 1962, 17, 29–40.

THIBAUT, J. W., & FAUCHEUX, C. The development of contractual norms in a bargaining situation under two types of stress. *Journal of Experimental Social Psychology,* 1965, 1, 89–102.

THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups.* New York: Wiley, 1959.

VALAVANIS, S. The resolution of conflict when utilities interact. *Journal of Conflict Resolution,* 1958, 2, 156–169.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received July 7, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 150-156

NO PLAY:
A MEANS OF CONFLICT RESOLUTION [1]

RALPH R. MILLER
Rutgers—The State University

In an iterative, individualistically motivated Prisoner's Dilemma type game played for monetary reinforcers, a no-play option was found to significantly increase the probability of resolving the conflict. By manipulating matrix values, 3 possible functions of the no-play option were factored out and examined: withdrawal, initiation of a change in strategy, and threat and punishment. Withdrawal was infrequently found to be the reason for a no-play choice. Threat and punishment were more often the explanation of the choice, but were ineffective in resolving the conflict. Initiation of change was frequently the motive for a no-play choice and was found to facilitate resolution. No sex differences in resolving the conflict were observed.

A. W. Tucher's Prisoner's Dilemma (PD) game,[2] shown in Figure 1, has frequently been employed as a laboratory model in the study of interpersonal conflict. Underlying its use is the assumption of a close analogy to a large class of "naturally occurring" human conflicts arising from nonparallel individual interests. There are, however, several other classes of "naturally occurring" conflicts similar in structure to the PD that have been excluded from consideration in PD discussions due to the occurrences of moves consisting of three or more possible choices.[3] This study examined a modified PD situation in which a third choice with an automatic trial payoff (overriding any choice by the other player) of zero-zero was available. Such a no-play, abstention, pass, etc., option may be considered alternately as part of the natural environment of the conflict or as a device added to an ordinary PD situation for the purpose of altering the probability of resolving the original conflict.

The singularly most consistent finding of the unmodified PD (Wilson & Bixenstine, 1961) is that most trials terminate with the mutually unsatisfactory conclusion, Element *22*, when the play is unitary or limitedly iterative (< 100). Deutsch (1958) has shown that the active cooperation necessary to resolve the conflict can be largely obtained by allowing the players to communicate. He hypothesizes that communications mediate the establishment of *trust*. Deutsch (1958) also found that permitting continuous choice reversal proved as effective an aid as direct verbal communications. The similarity of per-

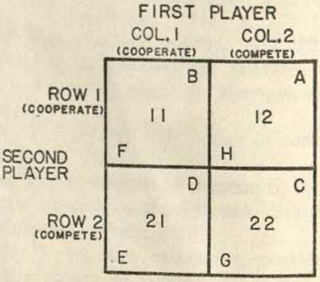


FIG. 1. Matrix for the unmodified Prisoner's Dilemma. (The first player chooses a column and the second player chooses a row. The letter in the upper right or lower left hand corner of the determined matrix element represents the payoff to the first or second player. The payoffs are subject to the constraints: $A = E > B = F > C = G > D = H$ and $B + F \geq A + H = D + E \geq C + G$.)

---

[1] This paper is based upon a master's thesis submitted at Rutgers University. The author is indebted to Jean Burton, William Pavlik, and Siegfried Streufert for their helpful comments. The research was supported in part by a Grant-in-Aid of Research from the Society of Sigma Xi and in part by United States Public Health Service Grant SFI-MH 311-02.

[2] The anecdote of the PD as well as an analysis and review of the typical PD situation created in the laboratory is readily available. For example, see Luce and Raiffa (1957) or Rapoport (1965).

[3] Using the terminology of Luce and Raiffa (1957), a *choice* is the alternative chosen at a decision point; a *move* is all the possible choices available at a decision point; and a *play* is the total sequence of choices which comprise the game.

mitted reversals to an iterative game implies that a series of moves may act as a communications device, albeit a meager one. This same idea was suggested by Luce and Raiffa (1957), and Wilson and Bixenstine (1961). In many situations outside the laboratory, iterative play conflicts must be resolved without verbal communications for establishing trust due to the players' mutual fear that any attempt at communications would weaken their own position. Frequently in these situations, a third choice corresponding to a zero-zero payoff is available. Although such a no-play type of choice is often made when permitted, its availability in many situations is overlooked.

A no-play choice can stem from three possible motives. First, it may be simply an "easy way out" of an Element 22 deadlock if Element 22 has a negative payoff. In this capacity, a no-play choice can be regarded as a withdrawal from the situation.[4] Secondly, if Element 22 has a positive payoff, a no-play choice by a potentially cooperative-acting player could punish and communicate a threat of future punishment to a competitive-acting player. Finally, a no-play choice might serve to communicate a desire to change strategies without threat to the other player or risk of loss to the user. This function is simply an applied form of Deutsch's "communication of trust" (1958; 1962), and could motivate a no-play choice regardless of Element 22 value. Switching from a competitive strategy to a no-play strategy "says something" to the other player, just as switching from a competitive strategy to a cooperative strategy would; only the former action avoids the risk of the latter choice.

The present study manipulated the Element 22 payoffs for the purpose of distinguishing the three possible functions of a no-play choice, withdrawal, threat, and initiation of change. For reasons already mentioned, initiation of change was expected to aid in resolving the conflict. Withdrawal presents an island of safety surrounded on both sides by risk; hence, it was expected to retard conflict resolution. This is consistent with a lack of communications being inherent in the withdrawal concept. Threat and punishment are known from both learning and conflict theory to be relatively ineffective in extinguishing behavior when no alternative choice of equal incentive is available (Estes, 1944; Schelling, 1960); hence, they were expected to serve as weak aids in resolving the conflict.

In addition to the no-play option availability and Element 22 value, a comparison between sexes was made to minimize the possible sources of within-groups error and to retest the usual finding (e.g., Bixenstine, Potash, & Wilson, 1963; Gallo, 1966) that sex is not a significant factor in PD type games. The sex differences reported by Rapoport and Chammah (1965), when compared to the numerous negative findings, appear to result from an enhancement of male cooperativity apparently produced by the unusually low payoff values. The present study used moderate payoffs of the order of $.05 per trial.

Wilson and Bixenstine (1961) also predicated that the communicative value of a third choice in a PD would aid in resolving the conflict, but failed to observe significant facilitation during free play. However, their third choice did not automatically produce a zero-zero payoff, for, if the other player made a cooperative choice, his payoff was of intermediate size in contrast to the zero payoff of the third choice player. This arrangement would discourage the use of the third choice by players who resent "giving something away for nothing," but if used would also carry a more obvious request for cooperation to the other player. Another possible explanation of their negative finding lies in the positive payoff values of Element 22. This would both lessen the desire to resolve the conflict and would add some element of punishment and threat to the communicative value of their third choice.

---

[4] As used here, withdrawal is a temporary condition that necessarily must be ended or reaffirmed on the next trial. This lack of permanence contrasts sharply with the more complete withdrawal of "taking my marbles and going home." While the latter type of withdrawal concludes the conflict in an obviously unsatisfactory manner, the former may be expected to have a less disastrous influence on the final conflict resolution in that it doesn't permanently close communication channels. (The threat of permanent withdrawal is worthy of future study.)

METHOD

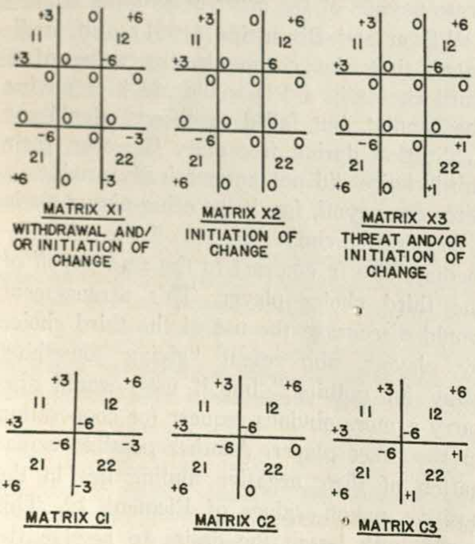
Subjects and Design

The subjects were 120 undergraduate volunteers from the introductory psychology classes of Rutgers University. Half of the subjects were females. The subjects were randomly paired by like sex, with care being taken that no subject knew the other member of the pair. Before the experimental session, the subjects knew only that the session would last about an hour and that they would have a chance to earn a small amount of money, the exact amount depending on how they performed.

A 2 × 3 × 2 factorial design was used corresponding to No-Play × Element 22 Value × Sex. The 60 pairs of subjects were randomly divided into six groups, each consisting of five male and five female pairs. The matrices are illustrated in Figure 2. The values shown are the actual payoffs in cents. Directly below the X matrices are the possible functions the no-play option could serve.

Apparatus

The male and female pairs used different but similar experimental rooms, each containing a room divider that prevented the subjects from seeing each other. Upon each subject's desk was an electronic apparatus, the top of which had a 3 × 3 matrix of red lights with both subjects' payoff values for each light directly below the light. Subject A (B) had three red choice buttons, one just to the right of (below) each row (column). Both subjects also possessed a green ready light and a black question button. The playing equipment was wired to the monitor in such a manner that the experimenter's turning on the ready light also unlocked the subjects' choice buttons until one was pressed. The monitor stored the choice until the other subject made his and then turned on the appropriate payoff light for each subject until the ready signal was sent out again. The center row and center column of lights, as well as the middle of the three choice buttons, could be covered, thus converting the 3 × 3 matrix into a 2 × 2 matrix.



MATRIX X1
WITHDRAWAL AND/OR INITIATION OF CHANGE

MATRIX X2
INITIATION OF CHANGE

MATRIX X3
THREAT AND/OR INITIATION OF CHANGE

MATRIX C1

MATRIX C2

MATRIX C3

FIG. 2. The six matrices. (The values shown are the actual payoffs in cents. Directly below the X matrices are the possible functions that the no-play option could serve.)

Procedure

The subjects were assigned to different waiting rooms and were individually escorted into the experimental room without seeing each other. Talking was prohibited and questions were to be written on the paper provided. Each subject was given a booklet consisting of the instructions, a test on instructions, the record form, and a questionnaire (postplay motivation test), respectively. As questions arose, the experimenter read them silently and then immediately answered them aloud. Fifty-six of the 60 pairs did in fact have questions, all of which were answered by direct readings from the instructions. When both subjects had finished the test on instructions, the tests were corrected and any errors that occurred were reviewed until there was no doubt that both subjects fully understood the instructions before the trials commenced.

To prevent last-trial profit taking, the subjects were told in the instructions that there would be between 20 and 40 trials rather than the actual number of 25. A trial consisted of each subject entering on the record form his expectation of the choice of the other player, what he thought the other player expected him to choose, his choice, and the reason for his choice in a brief sentence. A button was then pressed, and when both subjects had made their choice a payoff light went on for 15 seconds, after which the ready signal returned for the next trial. While the payoff light was on, each subject entered his earnings and actual choice of the other player on the particular trial. The subjects were all aware that the money was real and they would keep whatever they earned plus the $1.50 with which they started.

Care was taken to insure the most favorable conditions for the formation of an individualistic orientation. Separate waiting rooms prevented any interaction before the play. Notably, the subjects did not keep a record of the other player's earnings to prevent competition being instilled by continual comparison as found by Gallo, Irwin and Avery (1966). The subjects were addressed as "subjects" rather than "players," "experiment" replaced "game," and "earnings" replaced "winnings." The key phrase relevant to subject motivation in the instructions was "maximize your own earnings."

RESULTS AND DISCUSSION

The results of this study are presented in three sections, each focusing upon a separate measure of conflict resolution: the number of individual and pairs of cooperative choices, the individual and pair earnings, and pairs of cooperative choices by trial sequence.

Number of Cooperative Choices

The traditional dependent variable of the PD is the number of satisfactory conclusions, that is, Element 11 reached, because it is independent of the matrix values and hence allows comparison between the PD games using different matrices without introducing artificial differences. An analysis of variance of Element 11 conclusions found the no-play option to be significant ($F_{1, 48} = 5.0$, $p <$ .05). The value of Matrix Element 22, sex, and all interaction terms yielded *F*s of less than unity. The significance of the no-play option reflects the difference between a mean of 7.1 Element 11 conclusions with the option (28.4%), and 3.3 without the option (13.2%). This base rate of cooperative pairs is somewhat lower than usual but not unique (e.g., Wilson & Bixenstine, 1961), and is probably a product of instructional set. Due to this base rate, no conclusive verdict for a no-play option facilitation effect at high base rates can be drawn. The lack of significance between different Element 22 values implies that the no-play option served the same function under all three sets of values. Referring to the possible functions itemized in Figure 2, one can conclude that the no-play option acts primarily by communicating a desire to initiate a change in strategy, and that threat and withdrawal play no significant roles. The sex data, although insignificant, suggested less cooperativeness in the females consistent with all significant sex differences previously observed.

An analysis of individual choices yielded a significant no-play option ($F_{1, 108} = 5.6$, $p <$ .05) reflecting a mean of 10.5 individual choices to cooperate with the option (41.9%), and 7.9 without the option (31.6%). The differences between Element 22 values are also found to be significant ($F_{2, 108} = 4.1$, $p <$ .05), reflecting a mean of 10.2 cooperative choices with Type 1 (X1 and C1) matrices (41.1%), 10.2 with Type 2 (41.0%), and 7.0 with Type 3 (28.0%). Referring again to Figure 2, the difference appears to originate in an adverse effect of threat. Withdrawal still contributes little toward explaining the effectiveness of the no-play option. Considering the insignificance of threat in determining the number of Element 11 conclusions, threat apparently reduces the number of unilateral attempts at cooperation, but those that are still made are proportionally more effective in stimulating bilateral cooperation. Sex and all interactions were again found to be insignificant.

Earnings

The players' earnings are rarely used as a dependent variable due to the inherent problem of separating playing differences from artifacts created by the use of dissimilar sets of matrix values. Certainly this problem prevents a monetary analysis of the function of the no-play option as it would require the comparison of three matrices with different Element 22 values. However, there is considerable pragmatic value in examining the overall effectiveness of the no-play option in terms of dollars earned. The actual goal of each player is monetary profit, not Element 11 conclusions. Moreover, the argument against such an analysis is weakened as the Element 22 values of the no-play option condition are paralleled in the control condition.

An analysis of variance of individual earnings found the no-play option to be significant ($F_{1, 108} = 12.6$, $p < .001$), reflecting, after subtracting initial capital, the difference between a mean of \$.21$\frac{6}{10}$ with the option, and \$.04 without the option. An analysis of joint earnings by pairs yielded a significant no-play option ($F_{1, 48} = 10.7$, $p < .01$) reflecting adjusted pair means of \$.43$\frac{3}{10}$ with the option and \$.08$\frac{1}{10}$ without the option. The differefices between Element 22 matrix values were found to be significant at the .001 level in both cases. This of course must be attributed to the confound of design previously mentioned and cannot be construed to tell anything about the functioning of the no-play option. The *F* ratios for sex and all interactions were insignificant. In both the analysis by pairs and the analysis by individuals, the No-Play Option × Element 22 Value interaction displayed an insignificant trend toward a decrement in option effectiveness as

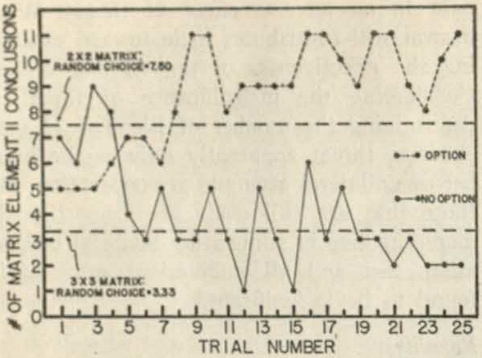


FIG. 3. Satisfactory conclusions by trials. (Both $N = 30$ pairs.)

element value increased. As this trend opposes the bias introduced by the diverse matrix values, it tends to support the previous choice analysis.

Sequence of Play

While both the number of cooperative choices and the dollar earnings are useful gross measures reflecting the overall outcome of a game, the trial-by-trial sequence of play offers a fine structure for analysis. Although innumerable questions can be asked of this data, the present analysis is limited to a brief examination of the ordered number of Matrix Element 11 conclusions.

The mean number of Element 11 conclusions by trials are presented in Figure 3. For purposes of comparison, the random choice expectations of Element 11 conclusions for the 2 × 2 and 3 × 3 matrices are also indicated. Most striking is the divergence of the groups with the no-play option and those without it, both from each other and their expectations given random choices. This difference, averaged over trials, was previously found to be significant. During the first six trials, the no-play option groups almost double their efficiency, while the control groups reduce theirs by one-half. The improvement of the option groups can probably be attributed to successful establishment of trust through communication of a desire to change strategies. The anomalous decrement of successful conclusions reached by the control groups has been observed frequently, and attempts to explain it have spoken of a "hardening of positions." More explicitly, the inability to communicate might prove frustrating (as was frequently reported after the experimental session), and the frustration (in this situation at least) might lead to aggression in the form of competitive playing.

The two curves of Figure 3 appear to be relatively stable after Trial 6. While not conclusive, the implication is that further trials beyond the twenty-fifth would have shown no major changes. One-hundred-plus trial games (e.g., Wilson & Bixenstine, 1961) usually have shown similar later-trial stability, although second-order trends have been observed in multihundred trial games (e.g., Rapoport & Chammah, 1965). This stability is, however, partly a statistical one, since six pairs of subjects who reached Element 11 at least five successive times saw fit to break the streak of satisfactory conclusions. These pairs were compensated for by new pairs "finding" the solution. It is doubtful, on the basis of this data, that either group would unanimously ever reach an endless streak of Element 11 conclusions. While group stability exceeded individual stability, individual stability also increased with trials. Between the second and twenty-fifth trials, the probability that a given subject would choose as he did on the last trial increased monotonically from 33% to 89%.

CONCLUSIONS

The no-play option is seen to have aided in resolving the conflict by serving as a device to initiate a change in strategies. Moreover, the threat function as predicted had no significant effect. The apparent failure of the withdrawal function to impair communications can be explained in that the subjects simply did not use the no-play option as an instrument for withdrawal, thus preventing a meaningful analysis of any potential withdrawal function. The subject's record form indicated his reason for each choice, and, incidentally, often reflected his interpretation of the other subject's last choice. Examination of these records for the subjects using matrix X1 (see Figure 2) found clear evidence of withdrawal in 7% of the no-play choices, and initiation of change in 27%. This contrasts with 31% of the no-play choices under X2 and 24% of the no-play choices under X3 being explained as initiation of change, and 30% of the no-play choices under X3 representing threat (and punishment). Threat, while present, was not effective, and withdrawal, while possibly effective, was not significantly present.

Using this same method of counting occurrences of reasons (done without knowledge of group assignment), a moderate confirmation of the no-play functions itemized in Figure 2 was possible. Threat was found as the reason for a no-play choice 14% of the time under X1, and 20% of the time under X2. Withdrawal was found as the cause of a no-play choice 2% of the time under X2, and not at all under X3. No other reason for the no-choice play (other than the proverbial "I felt like it") occurred with any regularity. Consequently, the choice of matrix values appear to have served rather well in factoring out the functions of the no-play choice except in the case of threat. While the trend still favors the predicted usage, threat is seen to be more readily used than expected. However, the range of threat usage, 14% to 30%, is at least large enough to have produced an indicative trend if threat had been effective as well as present. A final factor that would have enlarged any threat-mediated facilitation effect is that under X1 and X2 threat was rarely interpreted as such. All other functions were correctly interpreted about equally often, or 72% of the time an interpretation was offered (which in itself was only 14% of the time).

The generality of these conclusions must be limited to games played for moderate stakes ($.05 to $.10 maximum per trial), since Deutsch and Canavan[5] (1966) have recently found a significant decrease in the frequency of cooperative choices as the payoffs are increased from imaginary to moderate to large (≥ $1.00) values. Indicatively, four of the present subjects (all highly cooperative players) voluntarily reported after the experimental session that they would have played "more seriously" if the stakes had been higher. Gallo's (1966) conclusion of choice invariance in respect to payoff is limited by the relatively small value of his "large" payoffs. Certainly, further manipulation of payoff values in the future is desirable.

The problem of payoff size is only a part of the larger problem of motivation. After all efforts had been made in the instructions to produce an individualistic orientation, and this orientation was found to exist via the test on the instructions, no more could be done. However, after the game was concluded, a questionnaire given to recheck motivation found only 60% of the subjects to be still primarily individualistically motivated. Among the remaining subjects, 23% were found to be primarily competitively oriented and 17% primarily cooperatively oriented. No data was discarded due to the questionnaire. It was also found that 88% of the subjects would have preferred to talk to one another.

While many of the subjects played according to some system of strategy, there was a sizable fraction of the subjects, approximately 30% as determined from their stated reasons for choices, who were playing an almost random game. Many of these subjects professed having a "mental block" against any sort of arithmetic game and supported their self-evaluation by being slow in understanding the instructions. They were the players who most often used the "I felt like it" explanation of a choice. Clearly the problem is not just reaching a satisfactory solution to the dilemma, but for some subjects it is a matter of seeing the dilemma in the first place. It would be interesting to rerun this study (and others), using a subject population from which those who failed to comprehend the initial problem or failed to use something like logic in developing their strategy had been eliminated. On the other hand, if these people are in the majority or even a significant minority, perhaps conflict theory is using models less analogous to reality than is realized.

REFERENCES

BIXENSTINE, V. E., POSTASH, H. M., & WILSON, K. V. Effects of level of cooperative choice by the other player on choices in a Prisoner's Dilemma game: Part I. *Journal of Abnormal and Social Psychology,* 1963, 66, 308–313.

DEUTSCH, M. Trust and suspicion. *Journal of Conflict Resolution,* 1958, 2, 265–279.

---

[5] M. Deutsch & D. M. Canavan, personal communication, December 5, 1966.

DEUTSCH, M. Cooperation and trust: Some theoretical notes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1962*. Lincoln: University of Nebraska Press, 1962. Pp. 275–320.

ESTES, W. K. An experimental study of punishment. *Psychological Monographs*, 1944, **57**(3, Whole No. 263).

GALLO, P. S. The effects of modes of presentation and large rewards on a prisoner's game. Technical Report No. 18, 1966, University of California, Los Angeles, Contract Nonr 233(54), Office of Naval Research.

GALLO, P. S., IRWIN, R., & AVERY, G. The effect of score feedback and strategy of the other on cooperative behavior in a maximizing differences game. *Psychonomic Science*, 1966, **5**, 401–402.

LUCE, R. D., & RAIFFA, H. *Games and decisions.* New York: Wiley, 1957.

RAPOPORT, A. *The prisoner's dilemma.* Ann Arbor: University of Michigan Press, 1965.

RAPOPORT, A., & CHAMMAH, A. M. Sex differences in factors contributing to the level of cooperation in the Prisoner's Dilemma game. *Journal of Personality and Social Psychology,* 1965, 2, 831–838.

SCHELLING, T. C. *The strategy of conflict.* New York: Oxford, 1960.

WILSON, K. V., & BIXENSTINE, V. E. Effects of a third choice on behavior in a Prisoner's Dilemma game. Omaha: Nebraska Psychiatric Institute, 1961. (Mimeo)

(Received August 22, 1966)

Manuscripts Accepted for Publication in the *Journal of Personality and Social Psychology*

Behavioral Correlates of Academic Achievement: Conformity under Achievement- and Affiliation-Incentive Conditions: Robert S. Wyer, Jr.*: Department of Psychology, University of Illinois at Chicago Circle, Box 4348, Chicago, Illinois 60680.

Personality, Defensiveness, and Educational Achievement: Binyamin Z. Lebovits* and Adrian M. Ostfeld; Department of Preventive Medicine, University of Illinois Medical Center, 1853 West Polk Street, P.O. Box 6998, Chicago, Illinois 60680.

A Study of Social Sterotype of Body Image in Children: J. Robert Staffieri*: Pineland Hospital and Training Center, Box C, Pownal, Maine 04069.

Effectiveness of Debriefing following Deception Experiments: Elaine Walster,* Ellen Berschied, Darcy Abrahams, and Vera Aronson: Department of Psychology, The University of Rochester, River Campus Station, Rochester, New York 14627.

Immediacy, DRQ, and Content in Verbalizations about Positive and Negative Experiences: Roberta Gottlieb and Morton Wiener*: Department of Psychology, Clark University, Worchester, Massachusetts 01610.

Sex Differences in Types of Daydreams: Morton Wagman*: Student Counseling Service, 610 East John Street, University of Illinois, Champaign, Illinois 61820.

Personality Differences in Dissonance Resolution: O. J. Harvey* and Robert Ware: Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Role Playing: An Alternative to Deception: Martin S. Greenberg*: Department of Psychology, Washington University, St. Louis, Missouri 63130.

Interference, Immoderation, Inconsistency, and Dependency: Differences in the Behavior of Mothers toward First- and Later-Born Children: Irma Hilton*: 4 Washington Place—Room 666, New York University, New York, New York 10003.

Cooperative versus Competitive Concept Attainment as a Function of Sex and Stimulus Display: Patrick R. Laughlin* and Richard P. McGlynn: Loyola University, 6525 North Sheridan Road, Chicago, Illinois 60626.

Effect of Eye Contact and Physiognomy on Perceived Location of Other Person: Israel Nachshon and Seymour Wapner*: Department of Psychology, Clark University, Worchester, Massachusetts 01610.

Relationship of Success and Expectation to Affect on Goal-Seeking Tasks: Edwin A. Locke*: American Institutes for Research, 8555 Sixteenth Street, Silver Spring, Maryland 90910.

Relative Efficacy of Self-Monitored and Externally Imposed Reinforcement Systems: Albert Bandura* and Bernard Perloff: Department of Psychology, Stanford University, Stanford, California.

Independence and Conformity in Subsistence-Level Societies: J. W. Berry*: Department of Psychology, The University of Sydney, Sydney, New South Wales, Australia.

Communication of Affects between Feral and Socially Isolated Monkeys: Robert E. Miller,* William F. Caul, and I. Arthur Mirsky: Department of Clinical Science, University of Pittsburgh Medical School, 3611 O'Hara Street, Pittsburgh, Pennsylvania.

Voice Confrontation: A Bilingual Study: Philip S. Holzman,* Andrew Berger, and Clyde Rousey: The Menninger Foundation, Topeka, Kansas 66601.

* Asterisk indicates author for whom address is supplied.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 157–162

GROUP LEVELS OF ASPIRATION IN UNITED FUND CAMPAIGNS[1]

ALVIN ZANDER AND THEODORE NEWCOMB, JR.

Research Center for Group Dynamics, University of Michigan

The goals and changes in goals were examined in United Fund campaigns as examples of situations in which a subgroup must settle these matters for a larger organization. It was assumed that the product of the effort for which the goal was set was in itself valuable to all concerned for reasons beyond a desire to achieve success. It was found that a greater emphasis on increasing the income arose when the fund intake was smaller; this desire led to a repetition of the unattained goal, or to selection of a slightly higher goal, and to more frequent failures to attain the goal. Increases in level of the goal were followed by improvements in performance if the increases in the goal were not too large.

It often happens that a subgroup in an organization periodically must set a goal for the larger unit. The objective chosen in such a situation serves several functions: it defines a particular need of the inclusive body and it places a demand upon persons outside the goal-setting group. In performing its role, the subgroup is likely to receive influence attempts from parts of the larger body and to originate pressures itself on other parts. How does its unique position affect the goals the subgroup chooses? How do others react to these goals?

The official goals established for United Fund community campaigns, usually by a committee responsible for that decision, and the amount of money raised in the subsequent solicitations, provide data relevant to the phenomena of interest here. The present study is restricted to an examination of events in such financial canvasses during several years, yet its findings may be typical of organizations in which there are repeated demands on people to furnish other resources.

Level of aspiration theory offers a convenient model for approaching answers to the above questions in United Funds, since it considers the origins of repeated goals and their relationship to repeated performances (Atkinson & Feather, 1966; Lewin, Dembo, Festinger, & Sears, 1944). In studies of group-aspiration setting in the laboratory, moreover, it has been observed that the assumptions employed by members when selecting an aspiration level for their group are similar to the assumptions used by an individual when setting an aspiration for himself. A group aspiration, to illustrate, is chosen on the basis of the perceived probability of attaining the goal and the attractiveness of doing so; a more difficult goal is more attractive than an easier one; members raise the group's goal after the group has been successful and lower it after the group has failed; and group goals on the average tend to be slightly higher than past levels of performance (Zander & Medow, 1963). A stronger tendency to approach a given goal is aroused when members have a greater desire for achievement of group success (Zander & Medow, 1965; Zander & Wulff, 1966).

The official goals established by United Funds, and the changes in these goals, however, are not always what level of aspiration theory would lead one to expect. The apparent attractiveness of particular goals and the probabilities seemingly attributed to certain levels of performance, as we shall see, often are such that the chosen goals cannot be attained, yet they are seldom lowered—a desire for group success then does not appear to be the dominating motive in goal selection. Several conditions typically exist in a United Fund organization which can affect the probability and attractiveness that members of a goal-setting committee attribute to potential goals.

---

[1] This study was supported by the Air Force Office of Scientific Research, Contract No. AF49(638)-1630.

We are grateful for the help provided by Arthur Jette, United Community Funds and Councils of America and by John Forward, University of Michigan.

First, the official goal is a joint decision which interests many persons, inside and outside the goal-setting group. In selecting its goal the committee is exposed to pressures from citizens in the community, including those associated with the agencies supported by the fund, those responsible for conducting major phases of the campaign, as well as those whose support is essential to the success of the effort. External agents, it is known, can influence a group's choice of a goal and can do this even when the proposed goal deviates widely from the group's past level of performance (Zander & Medow, 1963; Zander, Medow, & Efron, 1965).

Second, those who are interested in the size of the fund, and who may seek to influence the official goal, place greater value on larger budgets. There is virtually no upper limit to the amount they may see as desirable, but there is some amount below which the budget should not be allowed to fall. An important consequence is that strong restraints exist against lowering the goal below what it has been in a prior year.

Third, goal setters assume that a community campaign can induce givers to provide what is asked. The purpose of a campaign goal is to arouse inclinations among potential contributors to give, and announcement of the goal is accompanied by social pressures exerted through advertising and solicitation which urge citizens to give what is needed.

Fourth, goals that contributors perceive to be unreasonably high have less influence on the size of the givers' contributions than goals they perceive to be necessary and attainable; a state of affairs goal setters may not always recognize.

These conditions suggest two assumptions about the nature of goal-setting in United Funds: (*a*) In choosing a goal, the decision-making committee is inclined toward selecting a higher goal than is reasonably attainable in the light of previous performance in that community; (*b*) this tendency becomes stronger as the need for more social welfare funds in the community becomes greater. The need becomes greater, it can be assumed, whenever a community fund fails to achieve its official goal.

Derivations from these assumptions describe specific events one may expect in the campaigns.

1. In general, a goal-setting committee whose official goal is not fulfilled is more likely to propose a more difficult future goal (for that community) than a committee whose goal is attained; it tries again to achieve at least the level of goal that was earlier failed. Thus, committees with more failing campaigns, compared to those with more successful campaigns (*a*) fix their future goals at a greater distance above past levels of performance, and (*b*) change the levels of their goals a smaller amount from one year to the next.

2. As the amount of money per capita solicited in a community is lower (in absolute terms) and the fund's need for more money is therefore greater, the discrepancy between the past level of performance and the level of the future goal is larger, not smaller.

3. The discrepancy between past performance and future goal generates less improvement of performance in failing funds than in successful ones.

METHOD

Each year United Fund organizations conduct financial campaigns in many communities for the support of social welfare programs. The campaigns are similar from town to town in their organizational structures and methods. The similarity in procedures is largely due to materials provided by the United Community Funds and Councils of America, a national association. A local fund typically announces a goal in advance and widely publicizes its progress toward that goal.

The national association publishes a *Directory* each year. This volume contains the name of each community holding a campaign, its population, the goal of the campaign, the amount obtained, and the amount raised in the immediately past canvass. The basic data for this study were taken from the directories for the 4 years from 1961 through 1964. Thus, we have goals for 4 years and levels of performance for 5 years. In addition, the same agency provides, for the last 3 of these years, a pamphlet entitled, "Determinants of Fund Raising Potential." This document reports the Effective Buying Income (EBI) in each of these towns and the "amount raised each year as per cent of EBI." The EBI is a summary index of financial data from each community, reported by *Sales Management Magazine*, describing the total amount of money available in the community for consumer purchases. The UCFCA suggests that the EBI index be used by campaign goal setters to allow a comparison of their local goal and performance with those of similar towns. Step-by-step procedures are provided for making these comparisons. It is not known how many funds do this, but probably more than a few do so. For present purposes the EBI will be taken as an indication of the amount of wealth in the community.

Data from communities providing complete information on all annual variables, save for a few omissions, were used. The towns chosen for this study had populations between 55,000 and 140,000, since the greatest number of cities reasonably similar in size was available within that range. The total number of communities within these restrictions, and the size of the sample, were 149.

From these records the following variables were employed: level of annual campaign goal, level of annual goal per capita, size of community, rate of community growth in population during the period observed, level of annual performance, and level of annual performance per capita. Two other variables were used which control on the wealth of the community. The *quality of performance* is the level of performance for a given year divided by the EBI for that year. This percentage, it should be noted, was available concerning every community in the country among the materials provided by the UCFCA. The *difficulty of the goal* is the level of the campaign goal for a given year divided by the EBI for that year. This percentage was not contained in materials provided by the UCFCA. The correlation between the level of campaign goal and its difficulty is .57 ($p < .001$); thus, we may assume that a higher campaign goal for a given community is usually a more difficult one.

RESULTS

Average Goal-Setting Events

It is informative to consider briefly the average values for separate events in the setting of United Fund goals. These results are reported on a per-capita basis for all communities in the sample and are the means of all four trials.

A fund establishes a particular *goal,* the number of dollars to be collected in the ensuing campaign. This is, on the average, \$3.49 per person. The campaign is run and the level of *performance* is \$3.34 per capita. Thus, the average campaign income differs from the average goal by an attainment discrepancy of \$−.15. A new goal is set and the average *shift in goal* level from the goal of the previous year is upward, \$.07. Thus the discrepancy between prior level of performance and the new goal (the *D* score) is \$.22. The primary tendency, then, is to raise the future goal even though the past performance failed to attain a lower goal. The average gift amounts to .165% of the citizen's personal EBI.

The relation of certain social characteristics in these communities to the goals and the levels of performance are of interest. These characteristics are: size of community, rate of population growth, and wealth (EBI) per capita.

Larger communities select higher goals per capita than do smaller towns ($r = .28$, $p < .01$), but the larger cities set less difficult goals, considering the EBI available to each person ($r = -.16$, $p < .05$). The size of the community bears no relationship to the quality of the performance (i.e., to the level of performance as percentage of EBI). The rate of community growth is not significantly related to difficulty of the goal or to the quality of performance, but faster growing towns more often raise their goals than do slowly growing ones ($p < .001$). Communities with greater wealth per person set higher per-capita goals ($r = .16$, $p < .05$) but these tend to be less difficult levels ($r = .31$, $p < .01$) and the quality of the performance is lower ($r = -.28$, $p < .01$) than in towns with less per-capita wealth. In summary, funds in smaller towns and in those with less wealth per person tend to have more difficult goals for citizens, and towns with less wealthy citizens show a better quality of performance per capita.

Derivations

In the first derivation it was proposed that committees for frequently failing funds, compared to those for more successful ones, tend to fix their future goals higher above levels of performance; that is, committees set larger *D* scores following failures than following successes.

In order to examine the relationship between *D* score and frequency of success or failure, it is most convenient to separate the 149 funds into sets, each having a different record of success during the years observed. This procedure provides five types of towns: those with four successes and no failures ($N = 27$), those with three successes and one failure ($N = 30$), those with two successes and two failures ($N = 28$), those with one success and three failures ($N = 37$), and those with no successes and four failures ($N = 29$).

The mean of the per-capita D score for 4 years for each of the community types was as follows: four successes–no failures, \$.06; three successes–one failure, \$.12; two successes–two failures, \$.21; one success–three failures, \$.25; and no successes–four failures, \$.42. The grand mean is \$.22. The distribution among these values is significantly different by median test ($\chi^2 = 46.35$, $p < .001$). The mean D score in the consistently failing towns is reliably larger than in the consistently successful ones ($t = 3.96$, $p < .001$). These results are in accord with the first derivation.

The towns with more frequent failures, it is worth noting, do not have higher goals per capita than those with more frequent successes; instead, as may be seen in Figure 1, they tend to have somewhat lower goals than the more successful funds. The difference in the level of the goals for the consistently failing and the consistently succeeding towns, however, is not a reliable one ($p < .20$). One can assume, then, that towns with different records of success are fairly similar in their goals.

In laboratory groups there is close adherence to the rule, "succeed, raise; fail, lower;" the level of aspiration is increased after a successful trial and lowered after an unsuccessful one. Do United Fund committees adhere to this rule? In four annual campaigns only three shifts in level of the goal are possible. Thus, in order to examine the goal shifts in communities with different records of success, during the campaigns for which these goals were selected the 149 towns were divided into four types: those with no successes and three failures ($N = 44$), those with one success and two failures ($N = 40$), those with two successes and one failure ($N = 33$), and those with three successes and no failures ($N = 32$). The dominant tendency in direction of goal shift was determined by an index in which each fund received a larger score when a shift in goal was upward from one year to the next and a smaller score when the shift was downward. The median index in the consistently failing towns was at the point indicating that no change in goal up or down was the most typical response; in the consistently successful towns the index was just below the maximum possible, indicating a strong tendency to shift goals upward. The distribution among the different index values was significant by median test ($\chi^2 = 10.82$, $p < .02$). These results again support the first derivation.

Also, the median amount that the goal was moved from the level of the previous year's goal increased as the frequency of success in the town increased. In communities with no successes and three failures the median amount of upward shift in the goal was \$.04, but in those with three successes and no failures the median amount of upward shift was \$.13. The trend between these two extremes was in regular order. The values in this distribution were significantly different by median test ($\chi^2 = 19.41$, $p < .001$).

Putting together the results concerning D scores and those describing shifts in goals from one year to the next, it is evident that the D scores are larger as shifts in goals are smaller. The correlation between these two is $-.27$ ($p < .01$). The larger D scores in failing towns, it appears, are not due to raising the goal levels but to persisting efforts to reach previously unattained objectives.

We earlier noted that the towns with different records of success do not have significantly different goals. Given the apparent restraints against moving the goals after a failing campaign, it appears that larger D scores more often occur when the amount of money raised is, in absolute terms, low. This is the content of the second derivation.

It can be seen in Figure 1 that towns with larger D scores (more frequent failures) solicit considerably less money than towns with smaller D scores (more frequent successes). This last difference is reliable ($t = 2.58$, $p < .02$). Larger D scores, then, are associated with lower levels of giving per person, and with the deprivation resulting from such giving. It is of incidental interest that the proportion of EBI provided per person in the failing towns is .136 and in the succeeding towns, .193. The direction of the difference suggests that frequently failing communities get a smaller proportion of each citizen's income, but the result is not statistically significant ($t = 1.27$, $p < .15$).

A quite different attempt was made to determine if a community with greater need for more social welfare funds sets a higher *D* score. In this instance a need for welfare funds was surmised from indirect measures. The relationship between the *D* score and the following demographic variables was computed within each of the five types of towns. The variables were: EBI per capita (assuming that a community with less wealth per citizen requires more social services), percentage loss of EBI during 4 years (same type of assumption), size of town (larger towns often have areas with greater needs for social welfare), and rate of growth in community population (rapidly growing towns have increasing demands). None of these variables was significantly related to the size of *D* score. These results also indicate, one should note, that towns with separate records of success or failure do not differ in these social characteristics.

The final derivation was that the amount of discrepancy between past performance and future goal generates an improvement in performance less often in failing communities than in successful ones. The data relevant to this derivation were obtained by relating the size of the *D* score in a given year and town with the amount of subsequent change in performance. In this analysis the immediately past level of performance serves as a common base for observing both changes in goal and changes in performance, while preserving the appropriate chronological sequence between them.

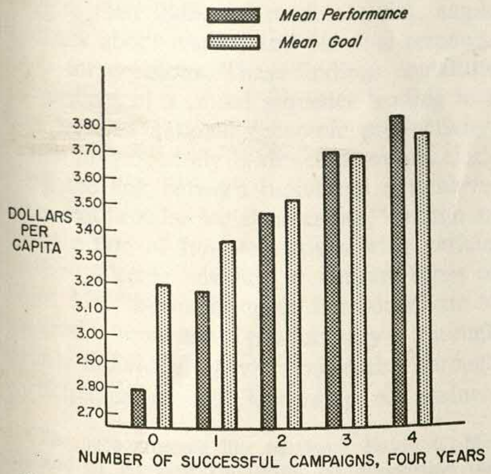


FIG. 1. Mean level of goal and performance in communities differing in number of successful campaigns.

TABLE 1

D SCORE AND SUBSEQUENT PERFORMANCE PER CAPITA

D score (goal for year $y + 1$)	No. campaigns	Change in performance (year y to $y + 1$)		
		Decrease or none	Moderate improvement	Much improvement
Negative	129	58%	36%	7%
Moderately positive	244	21%	53%	26%
Large positive	220	27%	27%	46%

Note.—$\chi^2 = 478.15$, $p < .001$.

Ignoring the successfulness of the communities for a moment, it is evident in Table 1 that performance tends to improve as the *D* score increases. The value of χ^2 for these distributions is unavoidably inflated since this analysis requires each community to be entered into the Table three times, once for each trial.

In communities with more successes than failures the correlation between a given *D* score and improvement in subsequent performance is .76 ($p < .001$), in towns with equal numbers of successes and failures the correlation is .73 ($p < .001$), and in towns with more failures than successes the correlation is $-.18$ (*ns*). It appears that a *D* score about the size set in failing towns (approximately $.40) is too large to serve as a stimulant to increased giving. These results support the third derivation.

An analysis of causation in the opposite direction, to determine if a specific amount of change in performance from one year to the next was followed by a comparable increase in *D* score, was not significant. The size of the *D* score appears to determine the subsequent change in performance, if any, rather than a given change in performance determining the change in subsequent *D* score.

DISCUSSION

The authors have examined the goals and the changes in these goals when a group must settle these matters for a larger social system.

The approach to this study was based on assumptions implicit in the theory of aspiration setting and was concerned with the conditions in a United Fund goal-setting committee which affect the perceived probability and attractiveness of a given goal. The results indicate that unreasonably difficult goals are more often chosen when performance in the local fund is deficient.

It is striking that a number of the funds fail and fail again but do not lower their goals to a level that will make failure unlikely in the future. Such results suggest that failure was not as intolerable to the goal setters as lowering the goal would be. One might conceive of the goal-setting committee members in the failing funds as having a fear of the consequences of failure. Groups in which such a fear has been induced among members have been observed elsewhere to set very difficult goals, apparently in that case because failure at a different task was less dissatisfying than failure at an easier level (Zander & Medow, 1965; Zander & Wulff, 1966). Given, however, that many communities were willing to repeat their failures, that the failures were not caused by higher goals in the unsuccessful communities than in the successful ones, and given the greater value placed on higher income, it seems unlikely that fear of the consequences of failure had much impact upon the goal-selecting process. Rather, the main reasons behind the high and unattainable goals in the failing funds appear to be (*a*) a need to fulfill large budgets which have not been met in the past, and (*b*) a belief that high goals and strong social pressures will arouse contributors to increase the size of their gifts. The results suggest that acting in terms of such beliefs is not always beneficial to a financial campaign.

REFERENCES

ATKINSON, J. W., & FEATHER, N. *A theory of achievement motivation.* New York: Wiley, 1966.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and behavior disorders.* Vol. 1. New York: Ronald Press, 1944. Pp. 333–378.

ZANDER, A., & MEDOW, H. Individual and group levels of aspiration. *Human Relations,* 1963, 16, 89–105.

ZANDER, A., & MEDOW, H. Strength of group and desire for attainable group aspirations. *Journal of Personality,* 1965, 33, 122–139.

ZANDER, A., MEDOW, H., & EFRON, R. Observers' expectations as determinants of group aspirations. *Human Relations,* 1965, 18, 273–287.

ZANDER, A., WULFF, D. Members' test anxiety and competence: Determinants of group's aspirations. *Journal of Personality,* 1966, 34, 55–70.

(Received May 31, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 163–168

THE ACHIEVEMENT MOTIVE AND ADVANCEMENT IN TWO TYPES OF ORGANIZATIONS

JOHN D. W. ANDREWS[1]
Brandeis University

This investigation deals with the business organization as a causal link in the demonstrated association between the level of a nation's achievement concern and its rate of economic development (McClelland, 1961). It is concerned with the joint effects of (a) motive strength of individuals, and (b) the evaluation standards of organizations, in determining who will advance most rapidly. It is hypothesized and shown that in a firm emphasizing achievement values (Firm A), individual advancement (as shown by frequency of promotions and raises and attained job status) is positively related to the strength of an executive's n Achievement. In a 2nd firm (Firm P) which values power striving, however, this relationship is reversed; in addition, n Power is negatively related to advancement in Firm A and positively related to advancement in Firm P.

Recently McClelland, in *The Achieving Society* (1961), has demonstrated a positive relationship between strong national achievement concerns, as coded from children's readers and other cultural products, and rapid economic development. He has also shown that this connection is mediated by socialization practices and adult occupational choice. That is, parents in achievement-oriented societies are more apt to value achievement and to use child-raising practices which are known to produce strong need for achievement (n Achievement). High n Achievement members of such societies, as they grow up, tend to choose business occupations as the most satisfying to their liking for moderate risk, ample feedback about results, and personal responsibility for decisions. These findings constitute the outlines of a causal sequence leading to a high level of national economic productivity.

The present study is directed toward clarifying the link between individual n Achievement and broader social processes, through an investigation of how executives with various motive patterns advance in various types of firms. Most businessmen do not contribute to national economic productivity through strictly individual activity, but rather through participation in and leadership of business organizations. One may extend McClelland's position to argue that a successful firm works well because it is organized to provide an effective channel for the economically productive efforts of its achievement-oriented members, and that a predominance of such firms should add up to a rapid rate of national economic growth.

It is well established that individuals with strong n Achievement work hard and effectively at tasks which provide the opportunity for attaining a personal standard of excellence (Atkinson, 1958, Ch. 19; McClelland, 1961, Ch. 6; McClelland, Atkinson, Clark, & Lowell, 1953, Ch. 8). Further, since businesses generally provide many opportunities for this sort of achievement, and benefit from it, it is to be expected that executives with high n Achievement will advance most rapidly (as indicated by promotions and raises) in their organizations.

This simple prediction is complicated, however, by the fact that such advancement is not solely a function of the individual's performance, but depends on the values and judgments of his superiors in the firm; and, while business in general may be favorable to achievement, it seems probable that firms vary somewhat in the singlemindedness with which they reward executive productiveness. This is particularly likely in a country such as Mexico, the site of the present study, where the dominant values differ as follows from the achievement-linked orientations of the United

---

[1] The author would like to thank David C. McClelland of Harvard University for supporting this research through a grant from the Carnegie Foundation, for his suggestions, and for the discussions that helped clarify the significance of the findings.

States: concern with "lineal" (older-younger) relationships, rather than with peer relationships or individual activity; preference for "being" rather than "doing"; fatalism rather than subjugation of nature; and present rather than future time orientation (Kluckhohn & Strodtbeck, 1961). These different value orientations will probably dictate in part the kinds of executive behaviors that are rewarded by advancement and thus the kinds of concerns it is useful for executives to develop in a given firm. Also, those who advance most rapidly are apt to be men who already share the firm's central values, and who should then use their increased influence in ways that will further strengthen those values.

The foregoing are the basic considerations that have generated these hypotheses:

1. When individual executives are tested for motivational concerns (such as n Achievement), the executive group of a given firm will have high scores on the motive that parallels the orientation of the firm as a whole.

2. Advancement in the firm will be a joint function of individuals' motive patterns, on the one hand, and dominant firm values, on the other. That is, rapid advancement (as measured by job status, promotions, and raises) is predicted in those cases where there is congruence between individual motive scores and dominant firm orientation. Less rapid advancement should occur when such congruence is absent.

METHOD

Firm Selection

The first step in the study was selection of two firms that would provide a clear-cut contrast in value orientations. It was specified that one firm should be highly achievement oriented, progressive and expansive in its policies, and economically successful. The other firm should be less achievement oriented, more conservative and traditional, and less successful economically. It became apparent very early that the second firm selected to be low on these criteria was strongly oriented toward power relationships. For this reason, the firms will be referred to subsequently as Firm A (achievement) and Firm P (power).

The selection of firms was made by Elliott Danzig,[2] an industrial psychologist working in Mexico, from among nine of his client organizations. He was provided with two descriptions of "pure" firm types, elaborations of the characterizations given above, and asked to match them as closely as possible. While his judgments are of an informal, nonquantitative type, they are based on several years' intimate acquaintance with the firms and are documented with specific observations; hence, they seem adequate for distinguishing the relative positions of the two organizations on the dimension specified.

Danzig's descriptions of the two firms are as follows:

Firm A. This is an American subsidiary, and prior to the study the parent company sent a new president (an American) to reorganize it. He gave the firm a thorough shaking up. New services and departments were instituted, including an Industrial Relations Department, Advance Planning Department, the launching of new products, etc. Production rose 40% in one of their four plants, actually to 115% of rated capacity, simply as a result of changes designed to make better use of managerial talent. The profits of the company began to rise about a year after the new president's arrival and, interrupted by the recent recession, have again hit record levels (1962). Productivity per employee has gone up. The company is doing well against its competitors in Mexico and is planning a move into a new area of production and marketing as well.

The success of the company is based in considerable part on a vigorous and continuing effort, initiated by the head of the firm, to get the right men in the right jobs. The new president undertook a thorough employee evaluation program as soon as he arrived, making extensive use of psychological testing and freely shifting men around so as to best use their capabilities. This man does not hesitate to advance a junior executive to high responsibility, if he seems able to handle it, or to fire an incompetent; his concern with evaluation is such that he has worked out his own point system for combining psychological evaluations into a single score which is used as a minimum criterion for consideration. He also shows considerable diplomatic skill in making men feel pleased with the changes that are made. He has actively urged the employment of Mexicans in high positions, even to the point of moving out Americans, and Mexicans can now look forward to just recognition of their capabilities.

The result of these policies is an atmosphere which provides many opportunities for achievement. Clearly visible to employees, most of whom are Mexican, is the fact that talent and effort are rewarded by increased position and opportunity to act. Many people respond to this challenge, and those who do not are replaced or advised of the fact and given help, if possible, through the counseling service. This company is a good example of the achievement-oriented business firm.

Firm P. This all-Mexican firm has a number of features which place it at the relatively nonentrepreneurial end of the scale. While in a relatively expansive area of economy, its early lead based on a large initial investment has been dwindling as the competition moves up. There has been only one expansion of operations in the past few years, partly because of discriminatory restrictions by the mayor of Mexico City. It has been argued by persons who know the firm and the city that a resourceful firm could find ways to circumvent such restrictions.

---

[2] Gratitude is due him for making available the data analyzed in this study, for his collaboration, and for his thoughtful suggestions about the investigation.

The internal environment of the company presents a picture consistent with the above. The firm is headed by a man who is generally very unpredictable. He runs his economic empire, playing the grand patron, rather like a feudalistic hacienda. All employees, good or bad, are considered the responsibility of the company, but an executive is quite likely to get a dressing down at one of the boss' command-performance gatherings for his staff.

This pattern has its effects throughout the organization. It produces insecurity and anxiety, an atmosphere in which an executive can never be sure of what the results of his decisions or attempts at creative ventures are going to be. This leads to timidity and a concern for hanging on to one's job rather than doing it well. Responsibility for getting a job done is not clearly given, and such is the fear of unpredictable consequences that few are willing to take the risk of making full use of the decision-making scope they do have. A further consequence of this is that people choose their subordinates with an eye to getting those who will not challenge their positions. They tend to prefer easily dominated yes-men who will not make their own work look poor by being too conspicuously productive.

Work in the company suffers from the effects of this entangling paternalism. There is reluctance to give rational, constructive criticism, with the result that employees get very little feedback about the quality of their work. Similarly, there is hesitation about firing an incompetent person, toward whom a paternal responsibility is apt to be felt. Thus Firm P shows few achievement-oriented characteristics and is more strongly oriented around issues of dominance and dependence.

In view of the strong dominance-submission or lineal power orientation present in Firm P, it was decided to study TAT-measured n Power as well as n Achievement, thus producing a symmetry that has made it possible to test the relationship between two types of motives and two types of firm orientations. The parallel between the description of Firm P and TAT themes scorable for n Power seems to have reasonably good face validity and is strengthened by the use of a stimulus picture that is very sensitive to superior-subordinate or lineal relationships (see discussion under Measurements section below).

Subjects

Subjects were selected as follows. All Mexican executives on whom complete data were available were included in the study, a sampling that reflects the nearly all Mexican composition of the firms. (Firm A had a few Americans.) Since the subjects were tested because of management referral for evaluation, they were probably not a random sample of the total executive group; their rate of promotion, for example, was somewhat higher in each firm than that for the total group. Also, there is no way of ruling out the possibility that executives were selected for testing by somewhat different means in the two films; one such difference, in proportions of upper- and lower-level men, has been controlled for in the data analysis pertaining to Hypothesis 1 (Tables 1 and 2). It seems unlikely, however, that different proportions of certain types of individuals in the samples from the two films could account for differences in the association between variables *within* a single firm, and, since it is these associations that are predicted in Hypothesis 2, it does not seem probable that they are artifactual products of such sampling variations.

Measurements

The variables discussed in the main hypotheses were operationalized via the following measures:

1. Individual strength of n Achievement and n Power was coded by the author in the standard manner (see manuals in Atkinson, 1958) from TAT stories obtained orally by clinical psychologists 2–4 years prior to the present study. All scoring was completed before the relation of other variables was examined. Three TAT pictures were used, the only three from a larger set which had been given to all subjects. These were: (*a*) a man in a white shirt sitting at what might be a draftsman's worktable with a small, unclear photo (possibly of his family) placed on it, (*b*) a nearly naked man climbing or hanging on a rope (Card 17BM of the Murray, 1943, series), (*c*) heads of two men, older and younger, close together ("father and son" card 7BM from Murray series). Virtually all of the power imagery was elicited by this last card, and this was entirely concerned with themes of fatherlike guidance, instruction, or domination of a younger man. This lineal power theme must be distinguished from other kinds of power situations (e.g., one in which peer influence is the theme) and is especially suitable for the purposes of this study. This is true, first, because of its relevance to the superior-subordinate relations which are very important in a business firm; and second, because of its congruence with the lineal, paternalistic pattern which is believed to characterize Firm P and Mexico in general as well (Kluckhohn & Strodtbeck, 1961; Lewis, 1961).

2. The variable of individual success in the firm was measured in three ways. The first measure was an estimate of job-status level at the time of the study; this was a measure of success since it presumably represents the outcome of an executive's efforts at advancement over an extended period. Job status was classified into two categories: Level I, major executives, decision makers in the firm; and Level II, minor executives, concerned primarily with decision implementation. These estimates were made on the basis of descriptions of the formal organization, such as organizational charts. It will be noted that since achievement-related behavior (taking moderate risks and responsibility, seeking feedback) centers on decision making and its conse-

TABLE 1

MEAN N ACHIEVEMENT SCORES CLASSIFIED ACCORDING TO FIRM (A AND P) AND JOB LEVEL (I AND II)

Firm	Job Level I			Job Level II		
	N	*M*	*SD*	*N*	*M*	*SD*
A	9	7.00	4.69	17	2.59	3.35
P	21	2.05	3.42	9	4.29	3.68
Difference (A − P)		+4.95[b]			−1.70[a]	

[a] $t = 1.16$, $p > .10$ (1-tailed test).
[b] $t = 2.86$, $p < .005$ (1-tailed test).

TABLE 2

MEAN N POWER SCORES CLASSIFIED ACCORDING TO FIRM (A AND P) AND JOB LEVEL (I AND II)

Firm	Job Level I			Job Level II		
	N	*M*	*SD*	*N*	*M*	*SD*
A	9	0.44	0.83	17	1.29	1.57
P	21	2.05	1.99	9	0.89	0.99
Difference (A − P)		−1.61[b]			+0.40[a]	

[a] $t = .79$, $p > .20$ (1-tailed test).
[b] $t = 3.13$, $p < .005$ (1-tailed test).

quences, Level I should contain all jobs for which n Achievement is a crucial personality characteristic. The second measure of success was the number of promotions received in the 4 years prior to the time of the study (1958–1962), and the third was the number of raises received in the same time period. Data on this last variable were, unfortunately, unavailable for Firm A.

It will be seen that firm values and evaluation criteria are considered to have been relatively constant over the 4-year period 1958–1962. Similarly, motive strength was measured as of the early part of that period (1958–1960) and is assumed to have been fairly stable over the same time span.[3] Thus the motive measures and the measures of firm orientation are the independent variables of this study, while the indexes of success, which are obtained either as totals over the 4-year period (promotions, raises) or at the end of the period (job status), are subsequent in time and are the dependent variables.

RESULTS[4]

Tables 1 and 2 are pertinent to testing Hypothesis 1; they present comparisons of mean n Achievement and n Power scores for the two firms, cross classified according to job level. An analysis of variance[5] performed on these data shows that overall firm differences (main row effects) are not statistically reliable, though they are in the right direction. The interaction effects are considerably stronger ($p < .10$), and comparison of the cell means indicates that for both n Achievement and n Power the firm differences are in the predicted direction and highly significant statistically, for the Level I subjects, with a much weaker trend in the opposite direction for the Level II group. In a sense, this finding anticipates the results concerning the relationship between motive strength and job level

TABLE 3

CORRELATIONS, BISERIAL (r_{bis}) AND PEARSONIAN (r), IN FIRMS A AND P BETWEEN MOTIVE SCORE (N ACHIEVEMENT AND N POWER) AND THREE MEASURES OF ADVANCEMENT (JOB LEVEL, FREQUENCY OF PROMOTIONS, FREQUENCY OF RAISES)

Motive	Job level: I or II[a]		Promoted: Yes or no[a]		No. raises	
	N	r_{bis}	*N*	r_{bis}	*N*	r
Firm A						
n Achievement[b]	26	+.64***	24	+.43***	21	+.36*
n Power[b]	26	−.39**	24	+.11	21	−.20
Firm P						
n Achievement[b]	30	−.37**	30	0.00		
n Power[b]	30	+.38**	30	+.34**		

[a] Used as dichotomous variable in calculation of r_{bis}.
[b] Used as continuous variable in calculation of r_{bis},
* $p < .10$ (1-tailed test).
** $p < .05$ (1-tailed test).
*** $p < .01$ (1-tailed test).

---

[3] Evidence supporting the assumption of motive stability is reported in an unpublished paper, "Patterns of Achievement Behavior in Business," written by the present author, which presents data on the retesting, in 1962, of about 67% of the subjects of the present study. The test-retest correlation (Pearson r) for n Achievement was .57 ($p < .02$) in Firm A, and, while the r was zero order in Firm P, a chi-square test, significant at $p < .05$, yielded a contingency coefficient of .37. For n Power the r was zero order in Firm A, and .33 ($p < .05$) in Firm P. Given that testing conditions were very different on the two occasions (oral versus written, individual versus group, assessment oriented versus research oriented, pictures versus brief stimulus phrases), these results indicate an adequate amount of motive stability; in addition, of course, the validity of the test scores, reported as the main results of this study, also support the assumption of motive stability.

[4] Since Hypotheses 1 and 2 both predict direction of difference and relationship, all significance tests are one-tailed.

[5] The analyses of variance based on the data presented in Tables 1 and 2 have been performed according to the approximation method suggested by Winer (1962) to take account of unequal cell frequencies.

and shows that the two questions cannot be completely separated. If we expect that individuals whose motives are congruent with the firm's orientation will gravitate toward the top, then it is appropriate to test the hypothesis concerning mean motive score differences between firms by focusing on executives who do reach the top, that is, the Level I group. In these terms Tables 1 and 2 confirm Hypothesis 1, since among Level I subjects Firm A is markedly higher in n Achievement and lower in n Power than Firm P.

Table 3 presents the relationship between motive strength and the three indexes of success, separately for Firms A and P. As predicted in Hypothesis 2, the correlations are significantly positive between n Achievement and all measures of success in Firm A, and between n Power and all measures of success in Firm P. Further, in three of five instances the nondominant motive is negatively correlated with measures of success, a finding which emphasizes even more strongly the differences between the two firms. In short, these data give firm support to the main hypotheses of the study.

DISCUSSION

The findings reported above show that the patterns of status and advancement in Firms A and P can reasonably be interpreted as the outcome of a mutual adjustment process between men with different motivational concerns, on the one hand, and firms with varying organizational values, on the other. In Firm A, the activities leading to job effectiveness and to advancement appear to coincide, a state of affairs which probably makes for economy of concern and effort among its executives. In Firm P, on the other hand, considerable conflict seems evident: while achievement concern may, as in Firm A, produce good job performance, it is, if anything, a negative factor in advancement; the latter requires a very different orientation to lineal power. In view of this tension between two basic goals of Firm P's executives, it is not surprising that the company as a whole should be less productive economically than Firm A.

The generally opposite relationships of the two motives to advancement in *both* companies emphasize the antagonism of the achievement and lineal power orientations per se.[6] This conclusion is similar to the findings (McClelland, 1961) that in the sphere of parent-child relations authoritarianism (especially on the father's part) inhibits n Achievement in the son, since such control is incompatible with the encouragement of independence that fosters achievement concerns. In organization life, too, it appears, the two orientations conflict and have sharply different effects as far as the encouragement of achievement is concerned.

These results, while confirming the importance of n Achievement for business activity, make it necessary to qualify the proposition that business is always achievement oriented and always attractive to high n Achievement individuals (McClelland, 1961). Clearly, this is much less true in Firm P than in Firm A. It seems likely that achievement-oriented businesses like Firm A will be more prevalent in high n Achievement nations, with the result that individual n Achievement is effectively channeled into economic productivity. In countries where other values predominate, however, business firms will probably partake of such values and be less purely achievement oriented (as is true of Firm P in Mexico); in such cases, whatever achievement concern exists in the country will be used ineffectively or deflected into other lines of activity,[7] and business will be less productive as a result.

Finally, it should be noted that the correlations between advancement and motive score (Table 3) constitute a refined and objectified version of the informal company descriptions with which the study began. Such correlations not only confirm the original judgments; they are also important new facts about the firms, being indexes of implicit or de facto organizational values. Such indexes, having been given preliminary construct validation in the present study, can subsequently be used in various new ways. One may, for example, compare them with what a firm's managers *say* they use as standards for advancement, or, again, with the "culture" of the firm, as coded from its documents in a way similar to that used by McClelland (1961) with children's readers at the national level. It should also be possible to group firms according to type of product, degree of economic success, nationality, or any other characteristic, and then compare the relations between personality and advancement that predominate within different groups. Finally, one might study changes in such correlations, for example, to assess the effects of a change in structure or in management policy.

---

[6] In this study, individual motive scores for n Achievement and n Power were also negatively related: Pearson $r = -.15$, $p < .15$.

[7] McClelland (1961, p. 264) suggests that this may be the case in Turkey, where middle managers have lower levels of n Achievement than school teachers.

REFERENCES

ATKINSON, J. W. *Motives in fantasy, action, and society*. Princeton, N. J.: Van Nostrand, 1958.

KLUCKHOHN, F., & STRODTBECK, F. *Variations in value orientations*. Evanston, Ill.: Row, Peterson, 1961.

LEWIS, O. *The children of Sanchez*. New York: Random House, 1961.

MCCLELLAND, D. C. *The achieving society*. Princeton, N. J.: Van Nostrand, 1961.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge: Harvard University Press, 1943.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received July 29, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 169–174

SEQUENTIAL EFFECTS IN SOCIAL INFLUENCE[1]

DAVID E. PETERSON,[2] HERBERT D. SALTZSTEIN, AND CHRISTOPHER EBBE[3]

Massachusetts Institute of Technology

Ss made numerical estimates of 8 displays of dots. Each display was judged 3 times. After each estimate S was exposed to a bogus estimate from a confederate peer. S's change in estimate within each block of 3 estimates, or "trial," constituted within-trial influence. S's change in estimate across the initial estimates of the blocks constituted between-trial influence. Between-trial influence effects occurred only when the confederate changed his estimates during each trial so as to maintain a given discrepancy from S's preceding estimate. No between-trial influence was obtained when the confederate maintained a fixed (discrepant) estimate throughout each trial. In contrast, neither of these 2 influence procedures produced any significant within-trial influence. Between-trial influence effects were also observed with judgments made in private.

Several years ago, Fisher and his colleagues demonstrated that social influence effects generalize to judgments of new objects. These investigators asked their subjects to estimate the numerosity of a given display, either an irregular pattern of dots (Fisher, Rubenstein, & Freeman, 1956) or a photograph of a large number of parachutists (Fisher & Lubin, 1958). Between the successive estimates of a given display (called a trial), they exposed their subjects to the estimates of a confederate which were consistently higher than those of the naive subject. Their data indicated that the influence communications not only effected a change in the estimates of the *same* display (*within-trial influence*), but also affected the *initial* estimate of the next display *prior* to any influence attempt on that new display (*between-trial influence*). That is, subjects who had been exposed to the influence message on a preceding display gave an initial estimate for the new display which was higher than estimates of subjects who had made their judgments in the absence of such prior influence attempts. In some cases, especially where the earlier influence messages were markedly discrepant from the judgments of the first display, the amount of overestimation of the second display was even greater than the change in the estimate of the previous display. Subsequently, Saltzstein, Rowe, and Greene (1966) also demonstrated between-trial influence effects in children using the same task but a somewhat different experimental procedure.

The major aim of the present study is to investigate how events within a trial affect carry-over between trials, specifically whether reduction of discrepancy within a trial enhances or reduces the between-trial influence effect. Fisher and Lubin observed that the amount of carry-over to a new trial was a positive function of the discrepancy (between the subject's and the confederate's estimates) on the previous trial. If this is generally the case, then a pattern of influence messages which permits reduction of discrepancy during the original trial should inhibit carry-over, while a pattern of influence messages which does not permit discrepancy reduction should enhance carry-over.

In order to test this hypothesis, the following two patterns of influence were introduced: In the first (*variable*) condition, each influence message is contingent on the immediately preceding estimate made by the subject for that display, and therefore changes as the subject changes his estimate within a trial, that is, during the successive estimates of that display. In the second (*constant*) condition, the influence messages the subject receives are contingent only on the subject's first estimate of a display and are constant within a trial. That is, in the second condition the subject is able to reduce the discrepancy between his estimate and the confederate's estimate within a trial by changing his estimate of the display under consideration. In the first condition this is not possible since, as the subject changes his estimate of the display, the bogus estimate he receives from the confederate also changes, thus maintaining a fixed distance between his immediately preceding estimate and the confederate's. Our first hypothesis was that between-trial influence would be greater in the variable than in the constant condition.

---

[1] This work was supported primarily by Grant MH 5249 from the National Institute of Mental Health, United States Public Health Service. Further aid was received from National Aeronautics and Space Administration Grant NsG 496 to Hans-Lukas Teuber, and National Institutes of Health Grant GM-1064-03 to David E. Peterson.

This paper is based in part on a thesis presented to the faculty of the Graduate School of Massachusetts Institute of Technology in candidacy for the MS degree by David E. Peterson.

[2] Now at the Pennsylvania State University.

[3] Now at the University of Missouri.

As previous writers have observed (e.g., Festinger, 1950), social influence interactions involve at least two kinds of issues. The first is the need to be accurate in making judgments, which leads a person to rely on information from others, especially when the object to be judged is presented in impoverished form, for example, very briefly. The second class of issues is more strictly social. Especially important social issues present in conformity situations are obligation (e.g., "group locomotion") and autonomy. The variable condition may be analyzed in these terms. Obligation is probably minimal since there is no goal interdependence, but we believe that the issue of autonomy is salient. From the subject's point of view, each successive estimate by the peer is not only a response to the stimulus, but also a response to the subject's previous judgment. That is, the confederate peer is perceived as not merely revising his estimate of the stimulus object (upward), but also moving away from (resisting or even rejecting) the subject's preceding judgment. Our second aim in this study was to separate these two factors. To do so we included the *delay* condition. In this condition the bogus estimates increased within a trial in the same manner as in the variable condition, except that the peer's estimates were received only at the end of a trial. Thus, the peer's estimates could not be interpreted as reactions to the subject's estimates, since they were presumably made by the peer without knowledge of the subject's own estimates. We assumed that influence procedures which aroused the subject's autonomy needs would elicit resistance to change. An interesting question was whether this factor would affect between-trial influence. Our second hypothesis was that there would be more between-trial influence in the delay condition—where the autonomy issue was minimized—than in the variable condition where it was prominent.

An additional control is necessary to test this proposition. In a condition in which the confederate reports all of his estimates at the end of a series of presentations of the same display, the subject may treat these reported estimates as a single estimate. Accordingly, a group must be run in which the confederate reports only a single estimate. This condition may be called the *one-influence* condition. Absence of a difference in influence between these two groups would indicate that the subjects in the delay condition were treating the group of estimates as a single estimate.

METHOD

Subjects

The subjects were 42 male undergraduates at the Massachusetts Institute of Technology. They were selected from an introductory psychology course which required participation in experiments. One subject's data were discarded due to a marked variability of his estimates on even a single display, and all analyses were based on the data from the remaining 41 subjects.

Procedure

A perceptual task involving the estimation of numerosity of dots was used. The display materials were eight slides, each display consisting of 119 dots arranged in an irregular pattern (actually one pattern of dots presented in each of the eight possible orientations). The subject made three estimates of a given display, once after each *presentation* of the same slide. This block of three estimates constituted a trial. The subjects were approximately 5 feet from a translucent screen, through which the displays were projected from the reverse side. The display on the screen was approximately 4 feet square. The slides were presented for 3 seconds, and the interval between any two successive presentations (including the interval between the third presentation on one trial and the first presentation on the next trial) was 15 seconds. The subject and confederate reported their estimates verbally, with the subject's estimates always coming first.

Unlike the earlier studies by Fisher and his coworkers where several displays differing in actual number of dots were used, a single display was employed throughout. The display was sufficiently irregular so that by presenting it in different orientations the subjects could be convinced that they were seeing several different displays. This method ruled out possible effects due to radically different sensory inputs which might obscure the social influence effects, of primary interest here.

Experimental Groups

Influence, variable (*I-V*). In this group ($N = 8$), estimates were reported after each presentation of each slide. The confederate's estimate was always 50% higher than the previous estimate of the subject. (I.e., the confederate's first estimate was 50% higher than the subject's first estimate, the confederate's second estimate was 50% higher than the subject's second estimate, etc.) Thus, throughout a block of three estimates (or trial) the confederate changed his estimate so as to maintain a given discrepancy from the subject's immediately previous estimate. No matter how the subject's estimate changed during a trial, this discrepancy remained the same.

Influence, constant (*I-C*). Estimates were also reported after each presentation of each slide in this group ($N = 8$), but in this condition the confederate's estimate was 50% higher than the subject's *first* estimate and remained constant throughout the trial (i.e., for the remaining two presentations). Thus, if the subject increased his estimate during a trial, the discrepancy between his and the confederate's estimate was decreased.

A comparison of Groups I-V and I-C and a suitable uninfluenced control group allows assessment of the effect of the two kinds of influence procedures on within- and between-trial influence.

Delay influence (*D*). In this condition ($N = 9$), the subjects and the confederate wrote down their estimates after each presentation of a slide, but did not report their estimates until the end of one trial. At that time the subject reported his three estimates first, and then the confederate reported his three bogus estimates. We assumed that this would reduce the salience of the autonomy issue. The confederate's estimates were 50% higher than each of the subject's estimates (as in the variable condition above).

One influence (*I-1*). In this condition ($N = 8$), the subject and confederate wrote down their estimates and, after the third presentation, reported only their third ("final") estimate. The confederate's estimate was 50% higher than that of the subject on the reported estimate.

A comparison between Groups I-V and D tests the effect of eliminating or reducing the autonomy issue on between-trial influence, with Group I-1 serving as the control for whether the subjects in Condition D treat the three estimates of the confederate separately or as one influence message.

Control (*C*). Subjects made their three estimates of each display in the absence of information concerning the estimates of others. In this condition, each subject was tested alone.

Instructions

The subjects were told that this was an experiment in perception, and that their task would be to estimate the number of dots which would be presented on the screen. The instructions were phrased to give the impression that the only reason the subjects were run two at a time was to save the experimenter's time.

At the end of the experimental session, the subjects were asked to fill out a questionnaire to determine whether or not they had guessed the purpose of the experiment. Some subjects guessed that one of the purposes was to study social interaction, but none suggested that the other subject was a confederate or that the displays were in fact one display.

Measures

Between-trial influence has been defined as the carry-over of social influence effects to a new item prior to exposure to judgments by another on that new item. Since the subject's *initial* estimate of any display (i.e., on any trial) precedes the confederate's estimate for that display, the rate of increase between these *initial* estimates may be taken as the measure of between-trial movement. A differential rate of increase between any of the experimental groups and the control group is evidence of between-trial influence.

The mean value of initial estimates for each group on each trial was computed, and a coefficient of regression of mean initial estimates on trials was calculated, together with confidence limits (at the .01 level). Any two groups were defined as significantly different at a given level of confidence if the confidence limits of the regression lines did not overlap at that level. This measure has the additional value of being independent of the starting point of the curve, that is, of the initial estimate of the first trial. Since there appeared to be an end

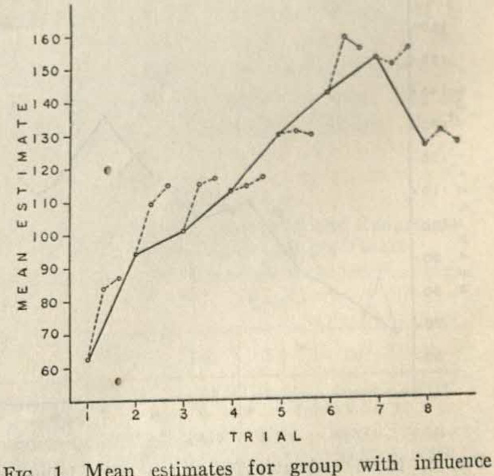


FIG. 1. Mean estimates for group with influence variable (I-V).

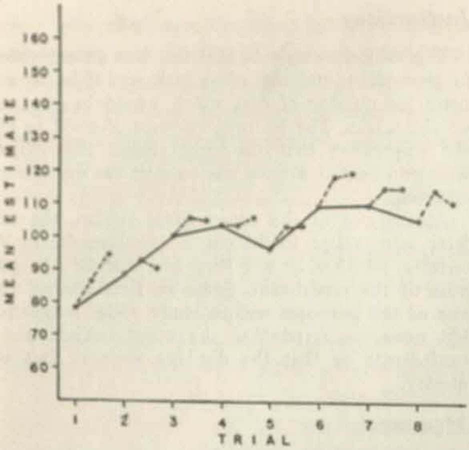


FIG. 2. Mean estimates for group with influence constant (I-C).

effect in the data, the last trial was not included in calculating the regression line.

The measure of within-trial movement was calculated as: (third estimate on trial n) − (first estimate on trial n). Each of the experimental conditions was compared with the control condition by the Mann-Whitney test. Significantly greater within-trial change in an experimental rather than the control condition would indicate that an influence effect was operating within the trials (or blocks of three estimates) in that experimental condition.

Both tails of the distribution were used in determining the significance levels except as noted.

RESULTS AND DISCUSSION

The first aim of this study was to discover the effect of two patterns of bogus estimates, one which changes throughout each trial so as to parallel the recipient's estimates (the variable condition) and one which remains constant throughout the trial (the constant condition), on between-trial and within-trial influence. The second aim was to determine whether reduction of the salience of the autonomy issue within trials by delaying receipt of the bogus estimates until the end of the trial enhances between-trial influence.

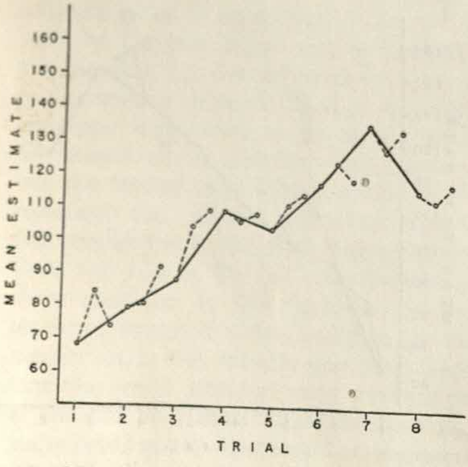


FIG. 3. Mean estimates for group with influence delayed (D).

The means for each of the 24 estimates (three for each display) are presented by condition in Figures 1–5.[4] Table 1 gives the regression coefficients on the initial estimates (across the first seven trials) and the associated confidence intervals for each group.

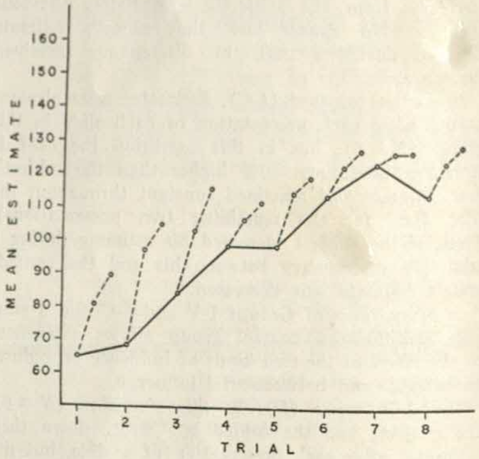


FIG. 4. Mean estimates for group with one influence (I-1).

These results indicate that the groups may be ordered in terms of between-trial influence effects as follows: variable > (one influence = delay) > (constant = control).

With regard to the first question, it appears that a pattern of bogus estimates which consistently increases within a trial promotes, and may even be necessary for eliciting, between-trial influence. This finding supports the observation by Fisher and Lubin (1958) that the carry-over effect is a positive function of the discrepancy between the subject and his

---

[4] The dotted lines connect estimates within each trial, within-trial movement; the solid lines connect the initial estimates on each trial, between-trial movement.

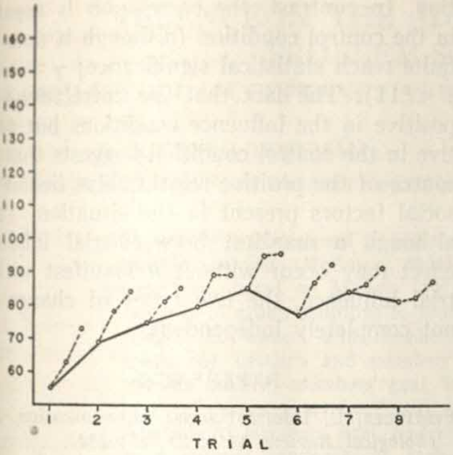


FIG. 5. Mean estimates for control group (C).

peer on the preceding trial. Our data have further detailed this observation by demonstrating that the discrepancy remaining at the end of the previous trial is a crucial factor. Contrary to Fisher and Lubin,[5] however, our data indicate no between-trial influence when the influence messages are constant within trials.

Since the one-influence and delay conditions do not differ in the amount of between-trial influence manifested, it is possible that the subjects were treating the three estimates of the confederate in the delay condition as a single estimate, and therefore no conclusions can be drawn about the second hypothesis. That is, the difference between the delay and variable conditions may not be due to the elimination of the autonomy issue in the former condition, but to the fact that the subjects were exposed to three influence messages in the variable condition, but effectively to only one in the delay condition. Nevertheless, the data suggest that, even if this was the case, the autonomy factor had a minimal effect.

The presence of between-trial influence in the one-influence condition is interesting, however, in another way. In that condition the subject reported only his third estimate in each trial, yet the between-trial measure is based on his initial estimate in each trial. The fact that the between-trial effect obtained for this group indicates that it is not necessary to require the subject to make a public judgment to obtain between-trial influence. Similarly, the lack of difference between the one-influence (private) and delay (public) conditions suggests that under the conditions of this experiment *between-trial* effects were not affected by the public versus private variable. This is consistent with the study by Saltzstein, Rowe, and Greene (1966), mentioned above, which demonstrated at least as much between-trial influence with children when judgments were made in a relatively private condition as in a public condition.

In marked contrast to the between-trial influence effects, within-trial influence occurs only in the one-influence condition, and there the difference from the control group reaches only a borderline ($p < .10$, one-tailed) level of significance. (Within-trial change tends to decrease over trials, in this condition, as it does in each of the other conditions.) It is clear from Figures 1–5 that subjects in the experimental conditions do vary and often increase their estimates within a trial, but so do subjects in the control condition. Typi-

TABLE 1

SLOPES AND CONFIDENCE LIMITS FOR THE REGRESSION OF MEAN INITIAL ESTIMATES ON TRIALS

	Group				
	I-V	I-C	I-1	D	C
Estimate increment	13.96	4.82	10.07	10.43	4.04
Confidence interval (.99)	13.65–14.27	4.03–5.61	9.79–10.35	10.02–10.84	3.10–4.98

---

[5] In the experiment by Fisher and Lubin, cited above, the influence messages did not change within a trial; that is, the procedure approximated the constant condition. The discrepancy between their findings and ours (in the constant condition) may be due to many factors. Two explanations, however, are especially plausible: (*a*) Subjects in their study were Army personnel tested in an Army setting; in our experiment subjects were Massachusetts Institute of Technology students tested in an university setting; (*b*) in their experiment subjects were given a warm-up period during which they were told to talk over any disagreements in judgments and try to be as accurate as possible. This treatment, absent from our study, may have enhanced the influence effect so that it appeared even when the influence messages were constant within trials.

cally the control subjects increase their estimates within a trial and then return towards the earlier estimates upon presentation of the new display. (This pattern of estimating in the control condition supports the assumption that the subjects were treating the displays as different in spite of the fact that they all contained the same number of dots.) The appearance of a between-trial influence effect in the absence of a within-trial effect is consistent with results reported by Saltzstein et al., mentioned above.

In order to explore further the relationship of within-trial to between-trial change, within-trial-change scores were summed across the first seven trials and correlated with the between-trial changes (i.e., with the initial estimate on the seventh trial − the initial estimate on the first trial) by Kendall's tau. The correlations in both experimental groups were positive, but reached an acceptable level of significance only in the variable condition where $\gamma = .71$ ($p < .02$) and in the constant condition where $\gamma = .62$ ($p < .05$). In the one-influence condition the correlation approached statistical significance ($\gamma = .55$, $p < .11$), and $\gamma = .18$ (*ns*) in the delay condition. In contrast, the correlation is negative in the control condition (although it does not quite reach statistical significance; $\gamma = -.52$, $p < .11$). The fact that the correlations are positive in the influence conditions but negative in the control condition suggests that the source of the positive relationships lies in the social factors present in the situation. Thus, although a manifest between-trial influence effect may occur without a manifest within-trial influence, the two types of change are not completely independent.

REFERENCES

Festinger, L. Informal social communication. *Psychological Review,* 1950, 57, 271–282.

Fisher, S., & Lubin, A. Distance and influence in two-person interaction. *Journal of Abnormal and Social Psychology,* 1958, 56, 230–238.

Fisher, S., Rubenstein, I., & Freeman, R. W. Interserial effects of immediate self-committal in a continuous social influence situation. *Journal of Abnormal and Social Psychology,* 1956, 52, 200–207.

Saltzstein, H. D., Rowe, P. B., & Greene, M. E. Spread of social influence on children's judgments of numerosity. *Journal of Personality and Social Psychology,* 1966, 3, 665–674.

(Received August 8, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 175–182

VALENCE IN THE ADOPTION OF SOLUTIONS BY PROBLEM-SOLVING GROUPS:

II. QUALITY AND ACCEPTANCE AS GOALS OF LEADERS AND MEMBERS[1]

L. RICHARD HOFFMAN AND NORMAN R. F. MAIER
Graduate School of Business, University of Chicago *University of Michigan*

This study was designed to test the generality of the valence model for group problem solving (Hoffman, 1961) for groups with leaders. 4 experimental conditions, based on Maier's (1952) distinction between the quality and acceptance of group solutions to problems, were established by offering monetary rewards to leaders and members separately for achieving their respective goals: (a) Leader's and members' goals, solution quality; (b) leaders' goal, solution quality; members' goal, leader's satisfaction; (c) leader's goal, members' satisfaction; members' goal, solution quality; (d) leader's goal, members' satisfaction; members' goal, leader's satisfaction. 9 3-man and 10 4-man groups were assigned randomly to each condition to solve a problem. The same relationships, shown in leaderless groups to support a model of valence in group problem solving (Hoffman & Maier, 1964), were uncovered in these groups with leaders and showed no variation across experimental conditions. The differential contributions of the leaders and members to maintaining this stability varied according to the goal instructions. Leaders tended to control the discussion by having the most influence over the final decision, regardless of their goals. The results point up the inadequacies of merely changing attitudes to elicit appropriate leadership behavior.

The present study was designed to extend our understanding of the relationship between the valence of solutions—their attractiveness or unattractiveness for the members—and their adoption by the group (Hoffman, 1961). In a prior study with leaderless groups (Hoffman & Maier, 1964), the valence of a solution, measured from a system of process observation, was shown to be positively correlated with the probability of its adoption by a group. Furthermore, the solution adopted by about 80% of the groups had the highest valence of all the solutions they discussed, and the adoption of this solution could be predicted at a relatively early stage of the problem-solving discussion. The use of this observational system in the present study permitted the authors to test the stability of these relationships and to examine possible variations of them in groups with leaders. In addition, the leaders and members in different groups, as a result of separate instructions, sought the same or different goals concerning the group's solution to the problem given them.

Two types of goals were used, consistent with Maier's (1952) distinction between the quality and acceptance of group solutions to a problem. Quality refers to the adequacy of the solution in terms of the objective facts of the problem. Acceptance refers to the members' motivation to carry out the decision. Therefore, in half the groups the leaders were given the objective of a high-quality solution (a monetary reward was promised), while the other leaders' objective was to gain the members' acceptance, measured by their rated satisfaction (Maier & Hoffman, 1960). Similarly, the members of half the groups received the high-quality goal, while the goal of the members of the remaining groups was the leader's acceptance. By instructing leaders and members separately, four combinations of leader's and members' goals were established. The principal question to be examined by this design was whether the acceptance orientation would reduce the tendency, especially among the leaders, for people concerned with solution quality to attempt unduly to influence the final decision. If merely changing the reward conditions could so influence these untrained leaders' behavior, the elaborate training programs in leadership skills (Maier, 1952) could be eliminated.

---

[1] The research reported here was supported by Grant MH-02704 from the United States Public Health Service.

METHOD

Problem

The Parasol Assembly Problem (Maier, 1952) was adapted from its role-playing form to a problem without roles. The problem involves a group of seven men who assemble carburetors in a series of sequential operations. An instance of inadequate production is described and the problem is to recommend the best means of achieving maximum productivity. A great variety of solutions are offered, but only about 10% of college-student groups usually develop the best solution. It should be noted that the problem appears to be intrinsically challenging to subjects and they usually spend the full 40 minutes available in trying to solve it.

Experimental Procedure

When the group members had convened, one person, preselected randomly, was told that he was the leader and was responsible for reporting the group's solution. The others were told that they were the members. A problem statement was given to each subject and then read aloud to them by the experimenter. The leader was then taken into the hall by the experimenter's assistant[2] and told that he would receive $1.50 if he achieved the goal set for him. For half the leaders the goal was a group solution of the highest quality (LQ condition), while the other half sought the members' satisfaction with the group's solution (LMs condition). At the same time the experimenter informed the members of their goal and of the reward of $1.50 each would receive if they achieved it.[3] The members of half of the groups had a group solution of highest quality as their goal (MQ condition), while members of the other half of the groups had as their goal the leader's satisfaction with the group's solution (MLs condition). In each group all members had the same goal. All subjects were told not to reveal their goals to the others in the group or they would forfeit the possible reward. No subject ever violated this instruction either directly or by implication.

Following the instructions the leader was returned to the group and the entire group was told to recommend the single best action to be taken to maximize productivity in the situation described. This instruction obviously complicated the task for the subjects with satisfaction orientations, since they had to appear to be working toward the quality objective in order to keep their instructions secret, while at the same time trying to gain their counterparts' satisfaction with the solution.

After the problem had been solved, each subject rated his satisfaction with the solution on a 6-point scale from "very dissatisfied" (1) to "very satisfied" (6).

Combinations of the two leader and member variations provided four experimental conditions: LQ-MQ, both leader and members have solution quality as their goal; LQ-MLs, leader seeks solution quality while members try for his satisfaction with the solution; LMs-MQ, leader seeks members' satisfaction with the solution, while members strive for solution quality; LMs-MLs, leader's goal is the members' satisfaction and the members seek his satisfaction with the solution.

Groups were assigned at random to the four conditions in a balanced order. Nine three-man groups and 10 four-man groups were assigned to each condition, a total of 76 groups.

Subjects

The 268 male subjects were obtained from the introductory psychology course as part of their class requirement and assigned as their schedules permitted to experimental groups. They were mostly freshmen and sophomores with only a vague acquaintance with the industrial scene.

The Observational System and the Measurement of Valence

The system used for recording the problem-solving statements has been described in detail in an earlier publication (Hoffman & Maier, 1964), and was modified only slightly in the present study. The observer records, for each statement subject makes about a particular solution, a number corresponding to the person and a letter indicating the category of statement made. There are five categories which indicate support for the solution (positive valence) and five corresponding categories of rejection (negative valence). Since each statement is recorded in the order it is made, a chronological record of the problem-solving process is obtained. No estimate of the length of each statement is made.

The valence of a solution for each individual in the group is measured by taking the algebraic sum of the number of positively valent statements he makes about a solution minus the number of negative statements he makes about it. The group valence measure is the algebraic sum of the individual valences for that solution.

---

[2] Bruce Springborn or the first author served as experimenter in all groups, and two undergraduates, James Ardis and Sheldon Shreiberg, served as assistants in giving instructions to the leaders and in coding the group process.

[3] The experimenter's assistant instructed the leaders while the experimenter instructed the members regarding the goals. It was found that when the experimenter's assistant instructed the members, subjects tended to regard the offer of reward as a psychological trick. As a result of this skepticism it was necessary to discard 13 groups. The problem did not arise when the experimenter instructed the members and the experimenter's assistant instructed the discussion leaders. The authors are quite disturbed about the image introductory psychology students have of the use of deception in psychological experiments.

TABLE 1

ANALYSIS OF VARIANCE DESIGN

Source	*df*
Leader goal (L)	1
Member goal (M)	1
Group size (S)	1
L × M	1
L × S	1
M × S	1
L × M × S	1
Error within cells[a]	68

[a] Used as error for all *F* tests.

Analysis Designs

The effects of the leaders' and members' orientations and of group size have been examined, for the most part, by either of two types of statistical techniques: analysis of variance or partition of chi-square. All metric data were subjected to the three-factor analysis of variance shown in Table 1, on the assumption that all variables were fixed. To conserve space the tabular presentation of results will indicate only those sources of variance which were significant at the .10 level. The *p* value will be shown in parentheses, for example, leader goal (.05). Frequency data were tested by partitions of chi-square which will be shown in the relevant tables.

RESULTS

Comparison with Study of Leaderless Groups

The principal objective of this experiment was to test the generality of the basic propositions of the valence model of group problem solving (Hoffman, 1961) for groups with formal leaders under the various experimental conditions. What effects, if any, did the experimental conditions have on the various relationships found earlier (Hoffman & Maier, 1964) between valence indexes[4] and the adoption of solutions by leaderless groups?

The first question to be raised is the issue of a minimum valence threshold for adoption, which was found to be about 14 in the leaderless groups. As in the previous study none of the groups in this study failed to submit a solution, but the lowest valence for a principal adopted solution was 8. It is not clear just how meaningful this number is, since the next lowest valence value for an adopted solution was 17. Certainly sampling variability around this threshold and the unreliability remaining in the coding system could produce a variation as large as this.

TABLE 2

MEAN VALENCE INDEXES FOR PRINCIPAL ADOPTED SOLUTIONS

Group size	LQ-MQ	LQ-MLs	LMs-MQ	LMs-MLs	Total
3 man	41.8	55.1	39.6	44.0	45.1
4 man	55.0	55.5	52.3	45.4	52.0
Total	48.7	55.3	46.3	44.7	48.8

Note.—No significant sources of variance.

More meaningful, perhaps, is the fact that the mean valence value attained by the principal adopted solutions in these 76 groups (48.8) was insignificantly different from that value in the 20 leaderless groups (51.4). Furthermore, although the range of mean valence values across conditions was from 39.6 in the LMs-MQ three-man groups to 55.5 in the LQ-MLs four-man groups, there was sufficient variability within conditions to produce no significant sources of variance when these were analyzed in the three-factor analysis of variance design (see Table 2). The valence value of the principal adopted solution in each group seems to have been peculiar to the unique character of the process in that group.

A second question concerns the frequency with which the groups adopted that solution, among the several discussed, which had the highest positive valence.[5] The principal adopted solution had the highest valence of all solutions discussed in 69 (90.8%) of the 76 groups and one of the subordinate adopted solutions had the highest valence in 3 additional groups. The 90.8% adoption rate in these groups is only slightly, and not significantly, higher than the 85.0% rate obtained in the 20 leaderless groups in the earlier study. Furthermore, the variation across conditions in the percentages of adoptions of the highest valent solutions—83% to 100%—was not significant. Thus the appointment of a leader

---

[4] See the Method section for a description of the valence index.

[5] Despite the instructions to all groups to submit a single action step for their solution, a number of groups proposed two or three steps as a single package. Prior to the analysis, the major part of such a package was identified as the principal adopted solution and the others were termed subordinate adopted solutions.

TABLE 3

RELATIONSHIP BETWEEN FINAL VALENCE INDEX AND ADOPTION

Valence index	LQ-MQ			LQ-MLs			LMs-MQ			LMs-MLs			3 man (N = 36 groups)			4 man (N = 20 groups)			Leaderless groups (N = 20 groups)		
	P[a]	S[b]	DNA[c]	P	S	DNA	P	S	DNA	P	S	DNA	P	S	DNA	P	S	DNA	P	S	DNA
40	11	0	0	14	2	0	11	2	2	11	0	1	19	0	2	28	4	1	10	0	1
30–39	6	1	2	4	1	2	5	1	3	4	0	0	8	2	5	11	1	2	7	1	3
20–29	0	3	4	1	2	4	3	1	1	3	2	4	7	5	5	0	3	8	1	3	5
10–19	2	6	8	1	3	14	1	0	12	1	2	14	3	5	23	2	6	25	3	4	11
1–9	0	0	103	0	1	90	0	1	90	1	1	94	0	2	172	1	1	205	0	3	111
0	0	0	72	0	1	53	0	0	68	0	0	45	0	0	96	0	1	142	0	0	83

[a] Adopted as principal solution.
[b] Adopted as subordinate solution.
[c] Solution discussed but not adopted.

and goal instructions had little effect on the tendency to adopt the solution with the highest valence.

The relationship between the valence level achieved by solutions and whether they were adopted as principal or subordinate solutions or not adopted at all was also tabulated. Table 3 shows these relationships for each of the four conditions, for the two group sizes, and, for purposes of comparison, the same relationship from the study of the leaderless groups. The similarity in the relationships is striking. The higher the valence value achieved by a solution, the greater was its probability of being adopted as the principal solution, regardless of goals, group size, or whether the groups had leaders or not. Chi-square comparisons across conditions and group size and between groups with and without leaders showed the valence level by adoption comparison to be the only significant one ($p < .001$).

Another finding from the earlier study was the location of an adoption threshold point with a valence value of about 15. In 88% of those leaderless groups the first solution whose valence equaled 15 was adopted as the principal solution. This analysis in the present study showed that 73% of the groups adopted the first solution whose valence equaled 15 (see Table 4). In another 4% of the groups the solution which first passed this threshold point was one of the subordinate adopted solutions. Although this percentage is less than in the leaderless groups, the difference is not statistically significant. Also, chi-square analysis showed no significant variations due to experimental conditions or to group size. Thus, the tendency was confirmed for a large proportion of groups to adopt the first solution for which the number of positive comments exceeded the number of negative comments by 15.

Finally, the earlier study revealed a correlation ($r = .59$, $p < .01$) between an individual's valence for the solution(s) adopted by his group and his rated satisfaction with the solution. The same correlation in the

TABLE 4

ADOPTIONS OF FIRST SOLUTIONS WITH VALENCE EQUAL TO 15

Group size	No. groups per condition	LQ-MQ		LQ-MLs		LMs-MQ		LMs-MLs		Total	
		N	%	N	%	N	%	N	%	N	%
3 man	9	5	62.5[a]	7	77.8	7	77.8	6	66.7	25	71.4
4 man	10	8	80.0	6	60.0	8	80.0	7	77.8[a]	29	74.4
Total	19	13	72.2	13	68.4	15	78.9	13	72.2	54	73.0

Note.—No significant comparisons by chi-square analysis.
[a] One group in this condition had no solution with valence greater than or equal to 15. The percentage is based on the reduced N as are totals which include the condition.

TABLE 5

CORRELATIONS BETWEEN INDIVIDUAL VALENCE FOR ADOPTED SOLUTIONS AND SATISFACTION WITH THE SOLUTIONS

	LQ-MQ	LQ-MLs	LMs-MQ	LMs-MLs	3 man	4 man	Total
Leaders	.51** (19)[a]	.24 (19)	.32 (19)	.35 (19)	.23 (36)	.46*** (40)	.36*** (76)
Members	.27 (48)	.34** (48)	.37** (48)	.19 (48)	.16 (72)	.38*** (120)	.28*** (192)
Total	.35*** (67)	.32*** (67)	.39*** (67)	.24** (67)	.20** (180)	.42*** (160)	.32*** (268)

[a] Numbers in parentheses are *N*s for correlations.
** $p < .05$.
*** $p < .01$.

present study, over all 268 subjects, was .32 ($p < .01$), which is significantly lower ($p < .01$) than in the leaderless groups. Whether this reduced correlation is a function, at least in part, of the varied goals of leaders and members could not be determined from the data. The two goals may have provided different bases for determining satisfaction.

The correlations between individual valence for and satisfaction with adopted solutions are shown for various sets of leaders and members in Table 5. In general, the correlations are all positive and of similar magnitude and, overall, about the same for leaders and members. Why they are generally smaller for the three-man groups is not clear, but suggests that the problem-solving process may have been somewhat different in the three-man and four-man groups. However, the variability among conditions may be due to the above-noted differences in goals.

Leaders' and Members' Contributions to the Valence of Adopted Solutions

Although it has been shown in Table 2 that the mean group valence index for the principal adopted solutions did not vary across conditions, one may still ask whether the leaders' and members' contributions to valence differed.

The data relevant to this question are presented in Table 6, where the leaders' and members' mean valence indexes for the solutions adopted by their groups are shown. The leader's valence for the adopted solution was affected not only by his own goal orientation, but also by the members' orientation. The LQ-MLs leaders had the highest mean valence, while LQ-MQ leaders had the lowest. The LMs leaders were not significantly different from each other, although the means of those with quality-oriented members tended to be slightly higher. Thus the LMs leaders were no less likely to influence the final group decision than were the LQ leaders.

The only significant source of variation for the members' valence was an interaction between the members' goal and group size. While in the four-man groups members with quality orientations were more influential than members with the goal of leader's satisfaction, the reverse was true for the three-man groups. There was no effect on the members of their leaders' orientations.

The data also permit us to look at the leader's contribution to the valence of the adopted solution in comparison with the average member's contribution. Was the acceptance goal effective in reducing the leader's tendency to dominate the discussion, thus equalizing the influence relationship between him and the members? For this analysis the mean valence index of the members was subtracted from the leader's valence index in

TABLE 6

MEAN VALENCE INDEXES FOR PRINCIPAL ADOPTED SOLUTIONS

Group Size	LQ-MQ	LQ-MLs	LMs-MQ	LMs-MLs	Total
Leaders[a]					
3 man	17.3	21.6	19.3	17.3	18.9
4 man	13.3	27.0	18.6	15.1	18.5
Total	15.2	24.4	18.9	16.2	18.7
Members[b]					
3 man	24.4	33.6	20.2	26.7	26.2
4 man	41.7	28.5	33.7	30.3	33.6
Total	33.5	30.9	27.3	28.6	30.1

[a] Significant effect: Leader Goal × Member Goal interaction (.05).
[b] Significant effects: Member Goal × Group Size interaction (.05), group size (.10).

TABLE 7

LEADER'S MINUS AVERAGE MEMBER'S VALENCE INDEXES FOR THE PRINCIPAL ADOPTED SOLUTION[a]

Group size	LQ-MQ	LQ-MLs	LMs-MQ	LMs-MLs	Total
3 man	5.0	4.7	9.2	4.2	5.8
4 man	−0.5	17.5	7.3	4.9	7.3
Total	2.1	11.4	8.2	4.6	6.6***

[a] Significant effects: Leader Goal × Member Goal interaction (.05), Member Goal × Group Size interaction (.10).
*** $p < .01$, by t test.

each group. The mean differences are reported in Table 7. In general, the leaders had significantly higher valence indexes, thus more influence (cf. Hoffman, Burke, & Maier, 1965), than did the average member by about 7 valence points (55%). When the data are examined by individuals (Table 8), we find that 52.0% of the leaders had the highest valence index in their group, as compared to a chance expectancy of 28.9% ($p < .001$). Thus the leaders, on the average, not only had more influence than the average member, but often had more influence than *any* of the other members.

The results in Tables 7 and 8 demonstrate also, however, some sharp distinctions among the four conditions (Leader Goal × Member Goal interaction, $p < .05$). LQ-MLs leaders were the most influential relative to the average member (by about 11 valence points), and most often (74%) had the highest valence for the adopted solutions. LQ-MQ leaders had about the same influence as the average member and were most influential with only about chance frequency (32%).

The most surprising result is the influence of the LMs leaders. In both LMs-MQ and LMs-MLs groups the leaders had higher valence indexes than the average member ($p < .01$) and had, significantly more often than chance, a higher valence than any member of their groups ($p < .05$). It would appear that in trying to achieve a solution satisfactory to all members, these leaders most often tried to persuade the members to adopt a solution that the leaders favored, rather than letting the members resolve their disagreements among themselves.

This is not to say that the members had no influence over the solution. A comparison of the grand means in Table 6 shows that the sums of the members' valence for the principal adopted solutions generally exceeded the leader's valence by about 11 valence points ($p < .01$). The members' totals exceeded the leaders' valence in all conditions, but espe-

TABLE 8

LEADERS WITH HIGHEST VALENCE FOR PRINCIPAL ADOPTED SOLUTIONS

Group size	No. groups per condition	LQ-MQ		LQ-MLs		LMs-MQ		LMs-MLs		Total	
		N	%	*N*	%	*N*	%	*N*	%	*N*	%
3 man	9	3	33.3	6	66.7	6	66.7	4.5[a]	50.0	19.5	54.2**
4 man	10	3	30.0	8	80.0	4	40.0	5	50.0	20	50.0***
Total	19	6	31.6	14	73.7***	10	52.6*	9.5	50.0**	39.5	52.0***

Summary of chi-square analysis (comparing leader highest and leader not highest)

Comparison	*df*	χ^2
Leader goal (L) × Highest valence (V)	1	0.01
Member goal (M) × V	1	2.97*
Group size (S) × V	1	0.13
L × M × V	1	3.81**
L × S × V	1	0.63
M × S × V	1	0.87
L × M × S × V	1	0.06
Total	7	8.48

[a] Leader was tied with another member for highest valence.
* $p < .10$.
** $p < .05$.
*** $p < .01$.

cially in the four-man LQ-MQ groups (mean difference = 28.4).

Thus, while groups in the various conditions did not differ generally in the valence they contributed to their principal adopted solutions, there were substantial differences in the leaders' and members' contributions to the valence. Where the leader's goal was quality and the members' goal was his satisfaction, the leader tended to contribute more valence and thus dominate the discussion. The leaders whose goal was members' satisfaction also tended to dominate their meetings, although not quite to the extent of the LQ-MLs leaders. Only in the LQ-MQ condition were the leaders generally undifferentiated from the members in their influence over the solution. In fact, in the four-man LQ-MQ groups the leaders' preference was often overridden by the combined efforts of several members.

DISCUSSION

The data are striking in the consistency of the relationships between the valence index and the adoption of solutions both between these groups with leaders and the leaderless groups reported on earlier (Hoffman & Maier, 1964) and across the experimental conditions, despite the substantial behavioral differences among leaders and members in the different conditions. Three such sets of consistent relationships were found:

1. The mean valence for the adopted solutions did not vary significantly across conditions nor between these groups and the leaderless groups.

2. The positive relationship between the valence index and the probability of solution adoption did not vary across conditions nor between these groups and the leaderless groups.

3. The percentage of adopted solutions which first reached a valence of 15 (73%), while slightly lower than the comparable percentage for leaderless groups, was not significantly lower, nor did it vary across conditions.

These relationships, when considered together, suggest (*a*) that a mechanism exists by which the group members "decide" that a solution which achieves a certain amount of support has been adopted, (*b*) that when this level of support has been achieved, resistance to that solution disappears or becomes relatively negligible, (*c*) that the discussion then focuses on providing further justification for adopting that solution, and (*d*) that the search for or the thorough discussion of additional alternatives is inhibited. Furthermore, the present evidence suggests that the level of support needed to activate this mechanism may be provided either by leaders or members, but that, once achieved, the leader takes an active role in directing the production of further positive valence for the solution.

To the extent that the leader is able to evoke positively valent comments from the members for the adopted solution, he may increase their satisfaction with it, as suggested by the positive correlation between the individual valence and satisfaction. However, the tendency among these leaders to provide their own justification for such solutions would seem to be reflected in their higher levels of satisfaction as compared to the members (leaders' mean satisfaction, 5.1; members' satisfaction, 4.7; p difference < .01). Leaders should be trained, then, in the skills of eliciting members' reactions to solution proposals if the latter's acceptance of adopted solutions is to be achieved.

The objective of gaining high-quality decisions also appears to run the risk of being defeated by this model problem-solving process. Since this process did not vary across conditions, the chance that a high-quality solution will be the first to surpass the adoption threshold would appear to be almost a matter of happenstance. Thus, one of the more important leadership functions, which must be developed through training, should be the prevention of solutions achieving this threshold before a sufficient number of solution possibilities have been brought forth for consideration by the group (cf. Maier & Solem, 1962).

In general, with respect to the effects of the different goals on the leaders' behaviors, these were largely overshadowed by their general tendency to dominate and control the discussion. Leaders exercised more influence on the adopted solution than did the average member of their group and, in more than half the groups, more than any other member, even though the leaders were assigned randomly.

This finding contrasts with the oft-reported relationship between personality characteristics and emergent leadership in leaderless groups (Mann, 1959, 1961; Hoffman, 1965). Thus, the goal of attaining members' satisfaction with the solution was insufficient to modify the behavior associated with the stereotyped role of the leader as one who dominates and controls the problem-solving discussion. Merely providing a reward for gaining the members' satisfaction is an inadequate substitute for training such relatively naive college students in the techniques for attaining this goal (Maier, 1963). Faced with the responsibility for producing a group solution and with the conflicting suggestions of the group members, these untrained leaders lacked the skills for resolving the conflict and chose, instead, to influence the group to adopt their (the leaders') preferred solutions.

REFERENCES

HOFFMAN, L. R. Conditions for creative problem solving. *Journal of Psychology,* 1961, 52, 429–444.

HOFFMAN, L. R. Group problem solving. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 2. New York: Academic Press, 1965. Pp. 99–132.

HOFFMAN, L. R., BURKE, R. J., & MAIER, N. R. F. Participation, influence, and satisfaction among members of problem-solving groups. *Psychological Reports,* 1965, **16**, 661–667.

HOFFMAN, L. R., & MAIER, N. R. F. Valence in the adoption of solutions by problem-solving groups: Concept, method, and results. *Journal of Abnormal and Social Psychology,* 1964, **69**, 264–271.

MAIER, N. R. F. *Principles of human relations.* New York: Wiley, 1952.

MAIER, N. R. F. *Problem-solving discussions and conferences: Leadership methods and skills.* New York: McGraw-Hill, 1963.

MAIER, N. R. F., & HOFFMAN, L. R. Quality of first and second solutions in group problem solving. *Journal of Applied Psychology,* 1960, **44**, 278–283.

MAIER, N. R. F., & SOLEM, A. R. Improving solutions by turning choice situations into problems. *Personnel Psychology,* 1962, 15, 151–157.

MANN, R. D. A review of the relationships between personality and performance in small groups. *Psychological Bulletin,* 1959, **56**, 241–270.

MANN, R. D. Dimensions of individual performance in small groups under task and social-emotional conditions. *Journal of Abnormal and Social Psychology,* 1961, **62**, 674–682.

(Received August 15, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 183–191

DEVELOPMENT AND EVALUATION OF AN ATTITUDES TO DEAFNESS SCALE

EMORY L. COWEN
University of Rochester

PHILIP H. BOBROVE
Veterans Administration Hospital, Montrose, New York

ALAN M. ROCKWAY
University of Rochester

AND JOHN STEVENSON[1]
University of Michigan

A network of 3 studies is reported, each relating to the development of an attitudes to deafness scale. The 1st of these, to develop a final 25-item test form, was based on face validity and item-test correlational techniques. In the 2nd study, interrelationships were demonstrated between antideafness and authoritarian, anti-Negro, and antiminority attitudes. These relationships were found not to be mediated by social desirability factors. The final study, using a "forming impressions of personality" paradigm, demonstrated that Ss with high antideafness scores, in comparison to their low-antideafness peers, gave more negative ratings to a confederate wearing a hearing aid than they did to a nondisabled but behaviorally comparable confederate.

The present report summarizes three investigations which have as their common objective the development of a measure of attitudes to deafness. Observers have pointed out that past attempts to collate views held toward various disability groups have been heavily biased by conjecture, impressionism, uncontrolled observation, and preconception (Barker, Wright, Meyerson, & Gonick, 1953; Levine, 1952; Meyerson, 1963; Wright, 1960). Levine (1952) has noted that such practice often leads differing observers to make directly contradictory pronouncements about a given group.

The need for reliable and valid objective instruments for assessment of attitudes toward disabling conditions is considerable. Such measures would help to establish normative attitude data for diverse groups. In addition they would provide a basis for understanding demographic and personalistic variables contributing to the formation of such attitudes and would facilitate evaluation of the effectiveness of procedures designed to modify attitudes. Specifically, with respect to the deaf, the need for such an attitude measure and for the types of research summarized above has been cited in a recent paper identifying research needs with this group (Rogers & Quigley, 1960). The development of an attitudes to deafness measure would expand considerably the potential both for basic and applied research in this area.

Recent work on a generalized measure of attitudes to disability has been reported by Siller and Chipman (1964). Attempts to develop attitude measures relevant to other specific disability groups have also been made. Reviews of such work for blindness and orthopedic conditions have been presented, respectively, by Lowenfeld (1963) and Cruickshank (1963). Several studies have been reported in which both specific and general attitudes toward many disability conditions have been compared (e.g., Gowman, 1957; Murphy, 1960; Murphy, Dickstein, & Dripps, 1960; Strong, 1931). In the main, these investigations seem to indicate that a generalized attitude of indifference is held toward the deaf, who are regarded in a less extremely negative manner than are several other disability groups (e.g., the blind). We are, however, still lacking a specific objective measure of attitudes towards auditory disability.

An earlier paper (Cowen, Underberg, & Verrillo, 1958) reported on the development and testing of a 30-item attitudes to blindness scale. An instrument with substantial internal

---

[1] Participation of the fourth author in the project was made possible through his appointment as a National Science Foundation Undergraduate Research Fellow.

consistency was developed, and it was found to correlate significantly with sociopsychological measures of authoritarianism, anti-Negro, and antiminority attitudes. In a subsequent investigation (Cowen & Cowen, 1964) an attempt was made to develop a crude attitudes to deafness measure to compare attitudes toward deafness and blindness within a French college sample as well as to study, cross-culturally, differences in attitudes to visual disability. That work was limited by the attitudes to deafness items which, rather than being created de novo, were adapted directly from the earlier attitudes to blindness items. The present research seeks to remedy that shortcoming and, through a series of interconnected studies, is designed to further the development of a new instrument to assess attitudes to deafness. The three component studies, targeted to this common objective, will be presented individually, followed by a unified overview. We shall first present the original test-development study including face validity and internal consistency data. Next, a correlational investigation will be reported, seeking to determine the degree of relationship between the attitudes to deafness scale and a series of other potentially related personality and attitude measures. This second investigation will include some cross-validational test-development data. And finally, there will be a predictive validity study designed to test the discriminating power of the new measure in a concrete situation which allows for prediction of behavior based on foreknowledge of the respondent's attitudes to deafness score. The behavior of concern is the relevant one of making judgments about the personalities of two people—one with normal hearing and the other ostensibly with auditory disability—following a brief question-and-answer-period interaction with them.

I. TEST DEVELOPMENT DATA

The 30 items making up the original attitudes to blindness scale (Cowen et al., 1958) were recast by substituting the terms "deafness" and "a deaf person" for "blindness" and a "blind person," respectively. Twenty additional items, each reflecting an attitude toward deafness or deaf people, were written based on statements about the deaf found in the literature. Often the latter included propositions without empirical base, reflecting little more than the prejudgments and stereotypes of a particular observer. The combined pool of items included both positively and negatively keyed ones (i.e., agreement with the given proposition indicating either a positive or negative attitude to deafness). The 20 new items were given to five clinically trained judges—two advanced clinical psychology graduate students, two clinical psychology PhDs, and one psychiatrist. The judges were asked to indicate whether agreement with each of the stated propositions reflected a positive or negative attitude to deafness or did not bear on the question of attitudes to deafness. For the 20 items so judged, there was 100% agreement on 18 and 80% agreement for the remaining 2.

The combined total of 50 items was then given to 100 introductory psychology students attending extension-school classes at the University of Rochester. Items were presented in a 4-point Likert-type rating framework (strongly or mildly agree or disagree) with no neutral point. Responses based on this administration were item analyzed according to Flanagan's (1939) method, the resultant item-test correlations ranging from 0 to .83. Twenty-five items, selected primarily on the basis of discriminating power, with item-test correlations ranging from .43 to .83, were chosen to constitute the final test form. These included 21 negative items (agreement indicates a negative attitude to deafness) and 4 positive items (agreement indicates a positive attitude to deafness). For scoring purposes, positive items were reverse keyed; hence, higher scores on the final scale indicate more negative attitudes to deafness. With 25 items and a maximum score of 4 points per item, the theoretical range of scores on the final form was from 25 to 100.

For purposes of the larger correlational study (cf. below) a second sample of 160 male introductory psychology students was also given the revised 25-item scale. This second administration provided a basis for cross-validating the initial set of item-test correlations. Table 1 presents the 25 final test items, keyed for direction of scoring, with item-test correlations reported for the two standardization samples. It is to be noted that the second set of *r*'s is somewhat lower than the initial set, ranging from .07 to .80. By and large, however, the shrinkage was relatively modest with only 4 items falling below an item-test *r* of .30 on the second administration. This amount of shrinkage is to be anticipated in terms of a probable element of capitalization on chance which entered into the final selection of items from the larger initial pool. Overall, there appears to be a reasonable stability to the discriminating power of the 25 constituent test items.

For the total sample, the mean antideafness score was 51.02 with a standard deviation of 7.32. In absolute terms, then, it may be said that the average score for the college sample as a whole had a central response tendency clustering around Point 2, thus reflecting a slightly positive attitude to deafness. Finally, a corrected split-half reliability of .83 was established for the final 25-item version of the scale.

II. CORRELATIONAL STUDY

Aim

The purpose of the correlational study was to examine the relationship between antideafness (AD) scale scores and a series of other attitude and personality measures which might have been expected, on some a priori grounds, to relate to this.

Measures

To this end, we used abbreviated versions of the Authoritarianism (F, 21 items), Anti-Minority (AM, 8 items), and Anti-Negro (AN, 12 items) scales from *The Authoritarian Personality* studies of the California group (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). In our earlier investigation (Cowen et al., 1958) these measures had been shown to relate significantly to an antiblindness measure, and it was hypothesized that a similar relationship might be anticipated when substituting other disability groups for the blind. The mediating theoretical notion was that the disabled in society share, with racial and religious minorities, an underprivileged status and, as such, some structurally common set of negative attitudes on the part of the nonhandicapped majority may be an-

TABLE 1

FINAL ATTITUDES TO DEAFNESS SCALE ITEMS WITH DIRECTION OF KEYING AND TWO INDEPENDENT SETS OF ITEM-TEST CORRELATIONS

Items	r_1[b]	r_2[b]
The deaf generally have a less mature personality than the hearing. (N)[a]	.64	.46
In general, deaf people are more neurotic than those who hear. (N)	.74	.36
It is impossible to really get "close" to a deaf person. (N)	.57	.59
Deaf people somehow seem sadder and more wrapped up in themselves than hearing people. (N)	.59	.80
The deaf do not seem to be bothered by ordinary life events any more than hearing people. (P)	.51	.33
Because of his need to be pitied, it is particularly important that the deaf person have someone very tolerant to whom he can talk. (N)	.49	.31
Deaf people also seem to have more than the usual number of *other* physical complaints. (N)	.46	.59
Deaf people show personality characteristics which frequently make them seem odd. (N)	.64	.66
A person who is deaf is as apt to be born a leader as anyone else. (P)	.46	.17
Deaf people seem to be overly polite and to lack spontaneity. (N)	.48	.79
Most deaf people feel that they are worthless. (N)	.59	.59
Most deaf people are dissatisfied with themselves. (N)	.53	.45
The deaf have as many interests as the hearing have. (P)	.52	.44
The deaf adult is not quite as mature or "grown-up" as the hearing adult. (N)	.56	.55
It's difficult to understand the deaf because they keep so much to themselves. (N)	.73	.62
It must be bitterly degrading for a deaf person to depend so much on others. (N)	.62	.39
On the whole, deaf children seem to be less intelligent than hearing children. (N)	.49	.51
I feel that deafness is as hard to bear as complete paralysis. (N)	.57	.32
A deaf person can't afford to talk back to people. (N)	.56	.07
You should not expect too much from a deaf person. (N)	.48	.62
A deaf person is constantly worried about what might happen to him. (N)	.60	.60
A deaf person is not afraid to express his feelings. (P)	.51	.15
Deaf people are more easily upset than people who can hear. (N)	.46	.12
The deaf are prone to have many more fears about the world than the hearing. (N)	.72	.60
The deaf are usually on their guard with people. (N)	.83	.58

---

[a] N indicates that agreement with the item reflects a negative attitude; P indicates that agreement with the item reflects a positive attitude.

[b] r_1 is based on 100 male and female introductory psychology students from evening extension-school classes; r_2 is based on 160 male introductory psychology students from the regular day-session classes.

ticipated. This viewpoint has been spelled out in detail for the blind under the title of the "minority parallel" by Chevigny and Braverman (1950), among others, and its extensibility to the deaf was subjected to test in the present study. The specific relationship between antideafness and authoritarianism might be anticipated on the basis of the anxiety the authoritarian person feels about weakness and defect. Such individuals may therefore stigmatize the disabled and form stereotypes about their supposed inadequacies and inferiority.

Conceivably, several different response styles could have mediated a series of positive intercorrelations among the AD scale and the three foregoing sociopsychological attitude scales. The agreement response set was in part controlled by the inclusion of reverse-keyed items on the AD scale. In order to assess possible effects of common social desirability response tendencies across tests, the 33-item Marlowe-Crowne Social Desirability (M-C *SD*) scale (Crowne & Marlowe, 1960) was used as a *control* measure.

On the assumption that a generalized predisposition to hostility might also relate to negative attitudes to deafness, 31 items from the Buss-Durkee Hostility Scale (BDHS) (Buss & Durkee, 1957) were added to the test battery. These included 10 items from the Irritability subscale, the intact 13-item Verbal Hostility subscale, and 8 items drawn from the Suspicion, Assault, Resentment, and Negativism subscales. For scoring purposes this yielded an Irritability, Verbal Hostility, and Total Hostility score, the latter overlapping considerably with the first two.

The last test included in the battery was the Locus of Evaluation scale (Rotter, Seeman, & Liverant, 1962) designed to measure internalization versus externalization of blame for a variety of events (I-E scale). It was thought that the individual who externalizes responsibility might be more likely to hold negative attitudes to the deaf than the person who is able to accept responsibility for things which go wrong.

Subjects and Administration

One hundred and sixty male white undergraduate students enrolled in an introductory psychology course were given the foregoing battery in group administration during the first week of the spring semester. The four Likert-type tests (F, AM, AN, AD) were intermixed in random order and presented in a common rating format. Following that, subjects responded to the M-C *SD* scale and BDHS items, both in true-false format and presented intermixed in a fixed random order. Finally, the 29-item I-E scale was given in paired-comparison format. The entire packet was stapled together in the order described above and presented in a series of large group-testing sessions lasting about 1½ hours each. The two experimenters for these sessions[2] were not connected in any way with the personnel involved in the predictive validity study to be described below.

Results and Discussion

Pearson product-moment *r*'s were computed, based on an N of 160, between AD scores and scores on each of the other scales in the battery. Three of these *r*'s were significant at $p < .01$ or beyond: F, AM, and AN, with values of .36, .50, and .50, respectively. None of the other correlations approached significance. Of particular interest is the complete absence of a relationship between AD and M-C *SD* ($r = .01$), suggesting that the positive correlations between AD and the three sociopsychological attitude indexes are not mediated by a generalized social desirability response stereotype. Rather, these three significant substantive relationships lend additional support to the view that the deaf are regarded in some ways as having attributes in common with other underprivileged minority groups in society and are negatively viewed by those who revere strength and authority. That such findings transcend any single disability group, such as the deaf, is suggested by the fact that they are almost identical both in magnitude and direction to earlier ones based on an attitudes to blindness scale (Cowen et al., 1958). Collectively from these two studies, we may infer that attitudes to disabling conditions share some generalized common elements with attitudes toward minority groups. Indeed it is conceivable that a general relationship of this type extends beyond minority groups. Supporting evidence for such a view is to be found in data reported by Gilbert and Levinson (1956) indicating a significant relationship between authoritarianism and rejection of mental patients.

---

[2] The authors express their appreciation to Robert Lee Carlisle and Gershen Kaufman for their assistance in this aspect of the data collection.

III. BEHAVIORAL PREDICTION STUDY

Introduction

Preliminary steps in the development of an attitudes to deafness scale appeared to have demonstrated reasonable face validity and internal consistency, and to have indicated a fairly stable pattern of interrelationships between this measure and several sociopsychological indexes of attitude. There remained, however, the question of whether responses on this scale, in fact, were predictive of some outside behavior, apart simply from responses to other attitude scales. It seemed desirable, in principle, to try to develop a criterion behavioral situation which, on logical grounds, might be expected to differentiate between subjects with high- and low-AD scale scores. This type of "acid" predictor test has been a chronic source of frustration in past work with disabled groups, since it is most readily approached through real-life situations which are frequently lacking in rigorous control and tend to yield data of a largely impressionistic nature. Illustratively, Gowman (1957) attempted to study certain attitudes to blindness by recording reactions of sales personnel in various stores to a supposedly blind customer who is seeking to purchase some shirts. Such a situation may provide excellent preliminary hunches and leads relevant to the problem of attitudes to disability, but it is extremely difficult to control the dimensions of interaction and to record and to quantify systematically the responses of "subjects."

For us, the theoretically important attributes of a task situation, to be used in a prediction of behavior study, were that it (*a*) have the qualities of plausibility and realism, (*b*) allow for the logical expectancy of differential behavior by high- as opposed to low-AD scale scorers on some relevant criterion, (*c*) have the potential for reasonably rigorous control in its implementation, and (*d*) be capable of yielding objective, quantifiable data.

It was decided that a "forming impressions of personality" paradigm based on triads including a "hearing" confederate and a confederate with an "auditory disability," in addition to the real subject, represented a model which gave maximal promise of meeting the foregoing criteria. After some substantial pilot work and debugging, a seemingly appropriate set of experimental operations was evolved.

That situation may be briefly described as follows: A real subject was to meet with the two stooges for a brief group-discussion period under the set of trying to find out as much as possible about the other two by questions and conversation. The structured rules of the game insured roughly equal time participation by each of the members of the triad. Prior "programming" of the stooges, as well as practice sessions, maximized the likelihood that they would be perceived as relatively neutral and roughly equal stimulus objects. Their participation as hearing and partially deaf stooges was counterbalanced. At the end of the conversation period, it was the task of the real subject to rate personal characteristics of the two stooges. Half of the selected subjects had high AD scores and half had low AD scores. Under these circumstances it was predicted that high-AD subjects would, in comparison to low-AD subjects, rate the auditorially "disabled" stooge significantly less favorably than the hearing stooge.

METHOD

Subjects

A total of 48 male subjects participated in the behavioral prediction study. These subjects were all drawn from a subject pool in introductory psychology, and all had been part of the sample of 160 used in the correlational study. Half of the selected subjects were extreme high AD scorers, and the other half were extreme low scorers. The respective cutting scores, based on the AD scale, for these two groups were 57 and above versus 45 and below. The correlational and behavioral prediction studies were separated in time by 2½ months, and they were run by completely different personnel. Hence, for all intents and purposes, they were entirely unrelated experiments—merely two of many conducted with introductory students during the semester.

ticipated. This viewpoint has been spelled out in detail for the blind under the title of the "minority parallel" by Chevigny and Braverman (1950), among others, and its extensibility to the deaf was subjected to test in the present study. The specific relationship between antideafness and authoritarianism might be anticipated on the basis of the anxiety the authoritarian person feels about weakness and defect. Such individuals may therefore stigmatize the disabled and form stereotypes about their supposed inadequacies and inferiority.

Conceivably, several different response styles could have mediated a series of positive intercorrelations among the AD scale and the three foregoing sociopsychological attitude scales. The agreement response set was in part controlled by the inclusion of reverse-keyed items on the AD scale. In order to assess possible effects of common social desirability response tendencies across tests, the 33-item Marlowe-Crowne Social Desirability (M-C *SD*) scale (Crowne & Marlowe, 1960) was used as a *control* measure.

On the assumption that a generalized predisposition to hostility might also relate to negative attitudes to deafness, 31 items from the Buss-Durkee Hostility Scale (BDHS) (Buss & Durkee, 1957) were added to the test battery. These included 10 items from the Irritability subscale, the intact 13-item Verbal Hostility subscale, and 8 items drawn from the Suspicion, Assault, Resentment, and Negativism subscales. For scoring purposes this yielded an Irritability, Verbal Hostility, and Total Hostility score, the latter overlapping considerably with the first two.

The last test included in the battery was the Locus of Evaluation scale (Rotter, Seeman, & Liverant, 1962) designed to measure internalization versus externalization of blame for a variety of events (I-E scale). It was thought that the individual who externalizes responsibility might be more likely to hold negative attitudes to the deaf than the person who is able to accept responsibility for things which go wrong.

Subjects and Administration

One hundred and sixty male white undergraduate students enrolled in an introductory psychology course were given the foregoing battery in group administration during the first week of the spring semester. The four Likert-type tests (F, AM, AN, AD) were intermixed in random order and presented in a common rating format. Following that, subjects responded to the M-C *SD* scale and BDHS items, both in true-false format and presented intermixed in a fixed random order. Finally, the 29-item I-E scale was given in paired-comparison format. The entire packet was stapled together in the order described above and presented in a series of large group-testing sessions lasting about 1½ hours each. The two experimenters for these sessions[2] were not connected in any way with the personnel involved in the predictive validity study to be described below.

Results and Discussion

Pearson product-moment *r*'s were computed, based on an *N* of 160, between AD scores and scores on each of the other scales in the battery. Three of these *r*'s were significant at $p < .01$ or beyond: F, AM, and AN, with values of .36, .50, and .50, respectively. None of the other correlations approached significance. Of particular interest is the complete absence of a relationship between AD and M-C *SD* ($r = .01$), suggesting that the positive correlations between AD and the three sociopsychological attitude indexes are not mediated by a generalized social desirability response stereotype. Rather, these three significant substantive relationships lend additional support to the view that the deaf are regarded in some ways as having attributes in common with other underprivileged minority groups in society and are negatively viewed by those who revere strength and authority. That such findings transcend any single disability group, such as the deaf, is suggested by the fact that they are almost identical both in magnitude and direction to earlier ones based on an attitudes to blindness scale (Cowen et al., 1958). Collectively from these two studies, we may infer that attitudes to disabling conditions share some generalized common elements with attitudes toward minority groups. Indeed it is conceivable that a general relationship of this type extends beyond minority groups. Supporting evidence for such a view is to be found in data reported by Gilbert and Levinson (1956) indicating a significant relationship between authoritarianism and rejection of mental patients.

---

[2] The authors express their appreciation to Robert Lee Carlisle and Gershen Kaufman for their assistance in this aspect of the data collection.

III. BEHAVIORAL PREDICTION STUDY

Introduction

Preliminary steps in the development of an attitudes to deafness scale appeared to have demonstrated reasonable face validity and internal consistency, and to have indicated a fairly stable pattern of interrelationships between this measure and several sociopsychological indexes of attitude. There remained, however, the question of whether responses on this scale, in fact, were predictive of some outside behavior, apart simply from responses to other attitude scales. It seemed desirable, in principle, to try to develop a criterion behavioral situation which, on logical grounds, might be expected to differentiate between subjects with high- and low-AD scale scores. This type of "acid" predictor test has been a chronic source of frustration in past work with disabled groups, since it is most readily approached through real-life situations which are frequently lacking in rigorous control and tend to yield data of a largely impressionistic nature. Illustratively, Gowman (1957) attempted to study certain attitudes to blindness by recording reactions of sales personnel in various stores to a supposedly blind customer who is seeking to purchase some shirts. Such a situation may provide excellent preliminary hunches and leads relevant to the problem of attitudes to disability, but it is extremely difficult to control the dimensions of interaction and to record and to quantify systematically the responses of "subjects."

For us, the theoretically important attributes of a task situation, to be used in a prediction of behavior study, were that it (*a*) have the qualities of plausibility and realism, (*b*) allow for the logical expectancy of differential behavior by high- as opposed to low-AD scale scorers on some relevant criterion, (*c*) have the potential for reasonably rigorous control in its implementation, and (*d*) be capable of yielding objective, quantifiable data.

It was decided that a "forming impressions of personality" paradigm based on triads including a "hearing" confederate and a confederate with an "auditory disability," in addition to the real subject, represented a model which gave maximal promise of meeting the foregoing criteria. After some substantial pilot work and debugging, a seemingly appropriate set of experimental operations was evolved.

That situation may be briefly described as follows: A real subject was to meet with the two stooges for a brief group-discussion period under the set of trying to find out as much as possible about the other two by questions and conversation. The structured rules of the game insured roughly equal time participation by each of the members of the triad. Prior "programming" of the stooges, as well as practice sessions, maximized the likelihood that they would be perceived as relatively neutral and roughly equal stimulus objects. Their participation as hearing and partially deaf stooges was counterbalanced. At the end of the conversation period, it was the task of the real subject to rate personal characteristics of the two stooges. Half of the selected subjects had high AD scores and half had low AD scores. Under these circumstances it was predicted that high-AD subjects would, in comparison to low-AD subjects, rate the auditorially "disabled" stooge significantly less favorably than the hearing stooge.

METHOD

Subjects

A total of 48 male subjects participated in the behavioral prediction study. These subjects were all drawn from a subject pool in introductory psychology, and all had been part of the sample of 160 used in the correlational study. Half of the selected subjects were extreme high AD scorers, and the other half were extreme low scorers. The respective cutting scores, based on the AD scale, for these two groups were 57 and above versus 45 and below. The correlational and behavioral prediction studies were separated in time by 2½ months, and they were run by completely different personnel. Hence, for all intents and purposes, they were entirely unrelated experiments—merely two of many conducted with introductory students during the semester.

Stooges

The entire procedure pivoted around exposure of a bona fide subject (extreme high *or* low AD) to a maximally controlled interaction with two other people who were to present themselves to the greatest extent possible as equally attractive human stimulus objects. The sole manipulated difference between these two confederates was that one was to be wearing a hearing aid and the other was not. To achieve these objectives, it was necessary first to articulate clearly the role of the stooges and then to practice until these roles were reasonably well perfected. The same two people, one a senior undergraduate psychology major and the other a first-year graduate student, served as the stooges for all subjects.

Several ground rules were followed in the programming and subsequent participation of the confederates. Each was in the hearing and auditorially disabled role in 50% of the cases, and, in each role, each took the position of first and second questioner in 50% of the instances. No matter what the role or order, each stooge asked the real subject the same number of questions. These questions were drawn from a preestablished list and tended to be relatively general and innocuous. (E.g., "What's your family background and what do your parents do?" "What are some of your interests and hobbies?" "What kinds of things have you done on your vacations?" "What kinds of jobs have you held?" "What kinds of things do you like to read when you have some free time?") The counterbalanced order of speaking of the two stooges insured that each posed these specific questions to the real subject approximately the same number of times.

In responding to questions posed by the real subject, the stooges were instructed to keep their answers relatively brief (15-60 seconds, depending on the question) and to try to adopt a relatively neutral stance. This meant providing relatively straightforward responses without excessive elaboration and trying to be neither particularly warm or friendly nor particularly hostile or threatening. As a kind of safety-valve precaution, we introduced a stratagem dubbed "bailing your partner out." We recognized that in the course of spontaneous interaction with real subjects whose behavior was *not* controlled our stooges might inadvertently be pushed into either a very friendly or very hostile reaction. Further, the real subject might (as a number did) pose questions which demanded that the stooge take some type of stand on a "loaded" issue. (I.e., "What do you guys think of what's going on in Viet Nam?" or "What should be done about drinking and sexual behavior in the dorms?") In such situations, it was almost certain that there was deep involvement by the real subject and that any substantive response by the first confederate would either endear or alienate him from the questioner. It was the responsibility of the second confederate to make his best clinical guess about what was happening and to assume a stance which was very similar to that of the first confederate—in other words, to counterbalance any unavoidable departure from neutrality by a response which would put him in the same position as his colleague. This was necessary to assure that, by the end of any given session, the two confederates would have approximately equal objective stimulus value for the real subject.

A practice period before the final experiment got underway was useful in identifying gross differences in the spontaneous interaction styles of the two stooges and in adding verbal correctives which made them more comparable as stimulus objects. All experimental sessions were observed, and continued feedback and correction were provided by the experimenter for the stooges to maximize the likelihood of their being similarly perceived. In the end, this key variable proved to have been quite well controlled, since the overall "level of liking" of the two stooges on the criterion evaluation measure was virtually identical. There were certainly no systematic preferences for one or the other of them, thereby lending some justification to the view of the presence or absence of hearing aid as the key independent variable.

Procedure

The actual subject and the two confederates were met by the experimenter in a waiting room outside of the experimental room. The experimenter greeted all three, took their names, and checked them off a master subject list. The three were then conducted to the experimental room, an observation chamber, with a table and several chairs in its center. A large microphone was in plain view, and on the table in front of each chair was an identifying card with a single letter, either A, B, or C, printed on it. Subjects were told to sit in any of the chairs, and no attempt was made to steer the real subject into any special position. It was, however, possible to control the positioning of the stooges so that each one was in the position of first or second questioner an equal number of times.

The experimenter then read aloud a set of printed instructions as follows:

> You are about to participate in a "forming impressions of personality" study. The situation is an extremely simple and straightforward one. You will have 20 minutes during which your object is to find out, by questions and conversation as much as you can about the other two people in the room with you. You are free to ask any question you wish, if you feel it may be helpful to you in forming your impressions. Questions about school work, hobbies, interests, background, family and future plans are all legitimate. These are just examples. You may wish to ask others. So that each person has an equal opportunity there are several simple ground rules that should be observed.
>
> 1) The three participants should take turns in asking questions in the order A, B, C.
>
> 2) When a participant asks a question it should be asked of *each* of the other two (A asks B and C; B asks C and A and C asks A and B).

> 3) Answers should be kept relatively brief (i.e., no filibustering) and each of you should try roughly to observe the rule of equal question-time and equal answer-time.
>
> After about 20 minutes I will return and ask each of you to rate the other two participants on a series of personality and behavioral dimensions. We are interested in comparing the accuracy of interpersonal ratings as a function of varying time acquaintanceship periods. We, therefore, hope that *Ss* will rate as accurately and frankly as possible under *all* rating conditions.
>
> I guess we can get started now. Just to get the ball rolling let me suggest an opening lead which we've found useful with previous groups. Why don't you each take a minute or so to let each other know what you're majoring in and some of the things you might wish to do when you finish up—in the order A, B, C.

At this point the experimenter left the room making it clear that he would be observing and listening to the subsequent discussion from behind the one-way-vision screen. The initial responses by the confederates to the experimenter's lead were broadly fixed, with allowance for leeway in specific detail. Each was programmed to respond in terms of any one of three or four "majors" and their logically related occupational choices. Each had preselected several specific major areas which, by virtue of background and interests, he felt competent to discuss as needed if probed further by the real subject. It was considered desirable to change cover stories to avoid any possible systematic liking or disliking for a particular major and to protect the credulity of the situation should there happen to develop any gossip among real subjects.

From that point on, the interrogation proceeded as per the instructions, with the confederates' questions coming in sequence from the pool of leads indicated above. The experimenter allowed the discussion to proceed through his own initial lead, plus two complete rounds of questions (A + B + C = 6 questions). Normally this required somewhere between 18–25 minutes. At the end of the second round, the experimenter returned to the room and announced, "Time's up now! We shall proceed to the evaluative portion of the experiment. I have some printed instructions here which I believe are quite clear. Read them over briefly to see if there are any questions."

The criterion instrument consisted of a semantic differential form on which the concepts to be rated were the two confederates as represented by their assigned letters (i.e., A and B, A and C, or B and C, depending on what letter the *true* subject had drawn). The actual rating task required locating the confederates (avoiding the use of tied placements) on each of the 10 semantic differential scales which were utilized for the evaluations: responsive-unresponsive, unpleasant-pleasant, mature-immature, withdrawn-outgoing, friendly-unfriendly, relaxed-tense, suspicious-trusting, cold-warm, interesting-dull, uncooperative-cooperative (Osgood, Suci, & Tannenbaum, 1957).

As soon as the real subject grasped the instructions, the three "subjects" were conducted to individual chambers where they were told they could do their ratings in privacy. The experimenter mentioned that he would like to speak with each individually for several minutes after they had finished their ratings and before they left. As soon as the doors to the rooms were closed the confederates left, while the real subject completed the semantic differential ratings—usually in about 5 minutes. The experimenter then interrogated him for several minutes in a relatively roundabout fashion, trying to assess possible awareness of the experimental operations. He asked about the insightfulness of the partner's questions, the validity of the experimental discussion-inquiry technique as a tool for finding out about others, and solicited suggestions for the improvement of the situation. Because of the total separation in form, personnel, and particularly time of the selection battery and the experimental test, it was virtually impossible for any subject to identify the exact purpose (or even substantive area) of the study and, indeed, none did. However, there were several subjects who were suspicious that their partners might have been confederates of the experimenter, having previously participated in an experiment of that type. These were eliminated and replaced with other subjects.

Results

Semantic differential ratings were converted into scores according to the following schema. For each subject and each pair of individual scale ratings, a minus was used to represent a less-favorable rating for the hearing-impaired confederate, and a plus represented a less-favorable rating for the normal-hearing confederate. The magnitude of the plus or minus was determined by the number of points separating the two absolute scale ratings (e.g., −2, +3, etc.). These item ratings were summed, retaining sign, across all 10 scale ratings, to yield a total rating score. Where the summed rating score was positive the hearing-impaired confederate was the preferred one, and when it was negative the hearing-impaired confederate was the less-preferred of the two. The mean summed rating score for the 24 low-AD subjects was +1.04, indicating virtually no difference in evaluation of the two confederates. For the 24 high-AD subjects the comparable mean score was −4.87, indicating directionally that these subjects preferred the hearing confederates to the one with the hearing aid. A *t* test of the significance of the difference between these two means yields a *t* value of 2.48, which, with $df = 46$, is significant at $p = .02$. Stated otherwise, subjects with relatively high initial AD test scores rated "subjects" with hearing aids significantly less favorably than their normal hearing peers, even though the overall stimulus attractability of these two types was controlled.

This analysis was repeated at the individual item level for each of the 10 constituent semantic differential scales making up the total test score. In all 10 instances the high-AD subjects gave relatively less favorable ratings to the hearing-aid confederate in comparison to the normal-hearing one than did the low-AD subjects. This difference reached statistical significance only in the case of the mature-immature scale ($p = .02$) and approached significance ($.05 > p > .10$) for relaxed-tense, suspicious-trusting, and uncooperative-cooperative.[3]

OVERVIEW

Each of the three individual studies presented in this paper represents a step in a comprehensive sequence designed to develop a useful measure for assessing attitudes to deafness. It seems appropriate, because of their logical interrelatedness and basically sequential nature, to discuss these three studies as a unit. The initial test-development study first established a pool of items which appeared, contentwise, to reflect attitudes to deafness. A judging procedure was utilized to establish some face validity for these items. Following a first round of responses by subjects to these items, it was possible to identify a subgroup of 25 test items with substantial internal consistency in terms of their relationship to total test score. A cross-validation study indicated that those initially selected items were relatively stable discriminators.

Next, a concurrent validity study, based on correlational procedures, indicated that negative attitudes to deafness were related to a cluster of sociopsychological attitudes including anti-Negro, antiminority, and proauthoritarian attitudes. The first two of these relationships may point up the tendency to view the disabled person as being in the same underprivileged position in society as the minority-group member; the third may reflect the inclination of the proauthoritarian individual to divide his world into the strong and the weak, to see the person with (auditory) disability as falling into the second of these camps which, for him, is a minimally respected one. Although the foregoing cluster of relationships seems quite clear-cut and consistent with earlier findings involving attitudes to blindness (Cowen et al., 1958), no relationships were found between attitudes to deafness and any of several measures of hostility and a measure of social desirability response tendencies. The latter is particularly interesting since it suggests that the significant relationships between the AD scale and the several sociopsychological indexes are not artifactually mediated by social desirability.

The utility of the prediction study resides in the demonstration that the discriminating power of the AD scale transcends paper-and-pencil inventories and extends to the world of real behavior. Negative attitudes to deafness as reflected item endorsements of respondents on this scale have been found, as might be anticipated on logical grounds, to relate to more negative ratings of a person with ostensible auditory disability in comparison to a normal-hearing individual. Since these differential ratings are made following a controlled triadic interaction, during which considerable pains have been taken to present the hearing and auditorially disabled subjects (confederates) as objectively equal person-objects in terms of their behaviors, verbalizations, and likability, we infer that they are, in part, a function of initial attitude. In this sense the scale gives evidence of predicting as it should in an experimental test situation which was set up so as to be sensitive to differential attitudes to auditory disability. Additional validating data is needed at this time, probing the discriminating power of the instrument in other settings where attitudes to auditory disability might be expected to be different either among different groups or as a function of experimental manipulations.

---

[3] These differences reach significance using a one-tailed test.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Barker, R. G., Wright, B. A., Meyerson, L., & Gonick, M. R. *Adjustment to physical handicap and illness*. (Rev. ed.) (Bulletin 55) New York: Social Science Research Council, 1953.

Buss, A. H., & Durkee, A. An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology,* 1957, **21**, 343–348.

Chevigny, H., & Braverman, S. *The adjustment of the blind*. New Haven: Yale University Press, 1950.

Cowen, E. L., & Cowen, R. S. Comparaison des attitudes d'étudiants Américains et Français à l'égard de la cécité et de la surdité. *Journal de Psychologie,* 1964, **61**, 201–208.

Cowen, E. L., Underberg, R. P., & Verrillo, R. T. The development and testing of an attitude to blindness scale. *Journal of Social Psychology,* 1958, **48**, 297–304.

Crowne, D. P., & Marlowe, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology,* 1960, **24**, 349–354.

Cruickshank, W. M. Psychological considerations with crippled children. In W. M. Cruickshank (Ed.), *Psychology of exceptional children and youth*. (2nd ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1963. Pp. 311–368.

Flanagan, J. General considerations in the selection of test items and a short method of estimating the product moment coefficient from the data at the tails of the distribution. *Journal of Educational Psychology,* 1939, **30**, 674–680.

Gilbert, D. C., & Levinson, D. J. Ideology, personality, and institutional policy in the mental hospital. *Journal of Abnormal and Social Psychology,* 1956, **53**, 263–271.

Gowman, A. G. *The war blind in American social structure*. New York: American Foundation for the Blind, 1957.

Levine, E. E. The deaf. In J. F. Garrett (Ed.), *Psychological aspects of physical disability*. (Office of Vocational Rehabilitation, Rehabilitation Service Series No. 210) Washington, D. C.: United States Government Printing Office, 1952. Pp. 125–146.

Lowenfeld, B. Psychological problems of children with impaired vision. In W. M. Cruickshank (Ed.), *Psychology of exceptional children and youth*. (2nd ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1963. Pp. 226–310.

Meyerson, L. Somatopsychology of physical disability. In W. M. Cruickshank (Ed.), *Psychology of exceptional children and youth*. (2nd ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1963. Pp. 1–52.

Murphy, A. T. Attitudes of educators toward the visually handicapped. *Sight Saving Review,* 1960, **30**, 157–161.

Murphy, A. T., Dickstein, J., & Dripps, E. Acceptance, rejection, and the hearing handicapped. *Volta Review,* 1960, **62**, 208–211.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

Rogers, M., & Quigley, S. P. Research needs in the vocational rehabilitation of the deaf. *American Annals of the Deaf,* 1960, **105**, 335–370.

Rotter, J. B., Seeman, M. R., & Liverant, S. Internal versus external control of reinforcement: A major variable in personality theory. In N. F. Washburne (Ed.), *Decisions, values, and groups*. Vol. 2. London: Pergamon Press, 1962. Pp. 473–516.

Siller, J., & Chipman, A. Factorial structure and correlates of attitudes toward disabled persons. *Educational and Psychological Measurement,* 1964, **24**, 831–840.

Strong, E. K. *Change of interests with age*. Stanford: Stanford University Press, 1931.

Wright, B. A. *Physical disability: A psychological approach*. New York: Harper, 1960.

(Received August 25, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 192–194

BRIEF ARTICLES

BIRTH ORDER AND DEPENDENCY[1]

RALPH H. STEWART

California State College at Los Angeles

The relationship between sibling position and dependent behavior was studied. 3 measures of dependency were used: Interpersonal Check List, Barron Items, and Embedded Figures Test. The 5 groups of Ss were oldest brother of brothers, oldest brother of sisters, youngest brother of brothers, youngest brother of sisters, and a "random" group. The latter group was not matched on the control variables size of family, presence of parents during childhood, or socioeconomic class, as were the 4 selected groups. The Embedded Figures Test was uniquely significant in that the oldest of brothers and sisters were more field dependent than were the youngest of brothers and sisters. The random group scored between these oldest and youngest groups.

Birth order as a behavioral determinant has a long research history with continuing interest seen in several recently published studies (e.g., Dember, 1964; Schachter, 1964). Systematic reviews (e.g., Warren, 1966) of the literature are available elsewhere and thus will not be presented here. The present study was designed to study the manner in which subjects representing four sibling positions differ on certain personality tests.

METHOD

Subjects and Procedure

Male students living in large dormitories at the University of California, Los Angeles, served as subjects. The students were asked individually if they would participate in a study which would take no more than 1 hour of their time. More than 99% of those approached consented, at which time they were asked to fill out a form to obtain information on the: (*a*) number of children in their family, (*b*) presence of parents during childhood, (*c*) sex of siblings, and (*d*) socioeconomic class (father's occupation). These forms took about 1 minute to complete, and each prospective subject was told that he would soon be contacted if he was among those persons finally sampled for the study. From these data forms, 5 groups of 20 subjects each were selected. Subjects were matched on each of the variables mentioned; that is, each subject came from a family of at least three children, had lived with both parents until at least age 15, and had a father working in the "white collar" occupations. Groups I–IV were composed of the following: I, oldest brother of brothers; II, youngest brother of brothers; III, oldest brother of sisters; IV, youngest brother of sisters. Group V was an unselected or "random" group; that is, the first four groups were matched on the variables described, while the fifth group was chosen at random from the original descriptive data forms. Following the selection of subjects, each subject was called upon in his dormitory room and administered three tests. The Interpersonal Check List (ICL) —La Forge and Suczek, 1955—was administered twice, first with instructions that subjects describe themselves as they think other people see them (Estimated Level I, Warren[2]) and then as they see themselves (self-perception or Level II). The paraphrased instructions included: "Check each adjective that describes you. For instance, if 'Assertive' applies to you, check it; if it does not apply, leave it blank."

Next, the Barron Items (Barron, 1953) were given with instructions to mark each item true or false depending on whether the subject agreed or disagreed with the statement. The Jackson (1956) abbreviation of the Embedded Figures Test (EFT) was administered. The standard procedure, as described by Witkin et al. (1954, 1962), was followed. The author administered all three tests to each subject.

RESULTS

The data were analyzed using a multivariate analysis of covariance as described by Cooley and Lohnes (1962). The first part of the analysis was a generalized analysis of variance. This test essentially pools all of the dependent variables with a significant result indicating that the groups differ in some way to be determined by subsequent analyses. Table 1 contains the results of the analyses with all five groups, that is, oldest of brothers (and sisters), youngest of brothers (and sisters), plus the random group. Table 2 contains the group means for all of the dependent variables.

---

[1] This paper is a portion of the research submitted in partial fulfillment of the requirements for the PhD degree at the University of Portland. The author wishes to thank the dissertation committee for their assistance: Walter G. Klopfer, chairman, Sheridan P. McCabe, Rolfe La Forge, and Carroll F. Raaum.

[2] S. Warren, "How Others See Us," unpublished manuscript, 1964.

The ICL scores need explanation. As seen in the Tables, Ain, Dom, and Lov are the ICL scores; Level I specifies "how others see you" and Level II, self-description. Ain is a score which measures the social desirability of the items checked. The fact that Ain I was significant indicates that social desirability was operative in the way the subjects responded. The dependency measure on the ICL was arrived at by obtaining the difference between the Love (Lov) score and the Dominance (Dom) score (see La Forge, 1963, or Leary, 1957, for scoring procedures). As Table 1 shows, the groups did not differ on the Dom and Lov scores. Thus, they do not differ on the ICL measure of dependency.

The EFT is uniquely significant, indicating that the groups differ in field dependence. Table 2 shows the group means. It can be seen that Groups I and III, oldest brother of brothers and oldest brother of sisters, took longer to find the embedded figures than did the other two groups. An unexpected result is that the random group scored between the groups of oldest versus youngest.

TABLE 1

MULTIVARIATE ANALYSIS OF COVARIANCE FOR THE EIGHT DEPENDENT VARIABLES: FIVE GROUPS

Test	*df*	*F*
All 8 variables	32,326	3.70**
Barron Items	4,95	0.81
EFT	4,95	17.32**
EFT 2 groups: oldest vs. youngest	1,78	69.56**
Ain Estimated Level I, Ain Level II	8,188	2.05*
Dom Estimated Level I, Lov Estimated Level I	8,188	0.74
Dom Level II, Lov Level II	8,188	0.53
Dom Estimated Level I, Lov Estimated Level I; Ain Estimated Level I as covariable		
Between groups	8,186	1.23
Covariable	2,93	16.58**
Dom Estimated Level I, Lov Estimated Level I; Ain Estimated Level I & No. children as covariables		
Between groups	8,184	1.22
Covariable Ain Estimated Level I	2,92	16.37**
Covariable no. children	2,92	0.01
Dom Estimated Level I, Lov Estimated Level I; Ain Level II, Dom Level II, Lov Level II as covariables		
Between groups	8,182	0.65
Covariable Ain Level II	2,91	1.32
Covariable Dom Level II	2,91	59.56**
Covariable Lov Level II	2,91	11.35**

* $p < .05$.
** $p < .001$.

TABLE 2

CELL MEANS OF THE VARIABLES: FIVE GROUPS

Variable	Group				
	I	II	III	IV	V
Barron Items	12.80	13.60	12.45	13.25	11.95
EFT	579.60	312.50	565.00	263.00	449.50
Ain Estimated Level I	2.14	2.04	2.03	2.08	2.18
Dom Estimated Level I	2.10	3.42	2.11	−0.54	−0.24
Lov Estimated Level I	−5.21	−4.19	−4.01	−0.70	0.72
Ain Level II	2.01	1.94	1.96	1.91	1.91
Dom Level II	−0.83	1.73	1.24	−1.28	0.79
Lov Level II	−3.04	−3.24	−2.62	−2.36	−1.40
No. children	3.25	3.30	3.20	3.25	3.05

DISCUSSION

The reason why these differences occurred only on the EFT is not clear. It may be that the other tests are susceptible to conscious manipulation and distortion (e.g., the significant ICL Ain). It has also been demonstrated (e.g., Mallory, 1965) that college students avoid dependency-connoting words. Another possibility which would need experimental clarification is that the interpersonal nature of the EFT affected the subjects differentially. The experimenter presented each EFT figure to the subjects. Although the experimenter was present in the room while the subjects filled out the paper-and-pencil tests, their interaction was presumably most apparent to the subjects during the EFT administration. Regardless of which arguments are used to "explain" this unique difference between the groups on the EFT, the possibility of such a result was considered while designing the study. Because of the difficulty in adequately defining dependency, it was questioned whether the instruments used would detect such behavior. The possibility of negative results raised doubts concerning their interpretability: would instrumental explication or the hypothesis be discredited? As is pointed out in any of the reviews on the subject, there is conflicting evidence as to the effects of ordinal position on dependency. This study has given further evidence of the ways in which such a relationship is evident.

REFERENCES

BARRON, F. Some personality correlates of independence of judgment. *Journal of Personality,* 1953, 21, 287–297.

COOLEY, F., & LOHNES, C. *Multivariate procedures in the behavioral sciences.* New York: Wiley, 1962.

DEMBER, W. N. Birth order and need affiliation. *Journal of Abnormal and Social Psychology,* 1964, 68, 555-557.

JACKSON, D. N. A short form of Witkin's Embedded Figures Test. *Journal of Abnormal and Social Psychology,* 1956, 53, 254-255.

LA FORGE, R. Research use of the ICL. Oregon Research Institute Technical Report No. 4. Eugene: Oregon Research Institute, 1963.

LA FORGE, R., & SUCZEK, R. F. The interpersonal dimension of personality: An Interpersonal Check List. *Journal of Personality,* 1955, 24, 94-112.

LEARY, T. *Interpersonal diagnosis of personality.* New York: Ronald Press, 1957.

MALLORY, G. The verbal conditioning of delinquent adolescents with neutral, dependent and aggressive stimuli. Unpublished doctoral dissertation, University of Southern California, 1965.

SCHACHTER, S. Birth order and sociometric choice. *Journal of Abnormal and Social Psychology,* 1964, 68, 453-456.

WARREN, J. R. Birth order and social behavior. *Psychological Bulletin,* 1966, 65, 38-49.

WITKIN, H. A., ET AL. *Personality through perception.* New York: Harper, 1954.

WITKIN, H. A., ET AL. *Psychological differentiation.* New York: Wiley, 1962.

(Received March 17, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 194-197

VARIATIONS IN MANIFEST ANXIETY DUE TO SEX, AGE,[1] AND EDUCATION

VASSO VASSILIOU, JAMES G. GEORGAS, AND GEORGE VASSILIOU

Athenian Institute of Anthropos, Athens

A Greek form of the Taylor Manifest Anxiety scale was given to a sample of 400 Athenian adults representative of the larger population with respect to sex, age, education, and income. Sex, age, and education were further analyzed using a 2 × 3 × 4 analysis of variance design. Sex and education were sources of variation. There was no interaction between sex, age, or education. Females gave higher anxiety scores than males. The higher the educational level, the lower the anxiety scores. The results suggest that sex and socioeconomic factors must be controlled in studies using the MA scale.

The Taylor Manifest Anxiety (*MA*) scale has been used to measure the role of drive or motivation in performance. The *MA* scale assumes ". . . that variations in drive level of the individual are related to the level of internal anxiety or emotionality, and . . . that the intensity of this anxiety could be ascertained by a paper and pencil test . . . [Taylor, 1953, p. 285]."

The *MA* scale has been given to various clinical and student groups. Taylor (1953) reports different distributions and median scores for students and psychiatric patients. Sex differences in students were found by Sinick (1956) and in children by Sarason et al. (1960).

The possible effect of sex and other socioeconomic factors on *MA* scale scores has not been systematically analyzed. Variations due to sex or socioeconomic factors, if not controlled, may obscure experimental results. Before the *MA* scale can be used in experimental studies, the variations of scores in the general population should be known.

This study investigates the effect of sex, age, and education on *MA* scale scores in an urban population.

METHOD

The 50 items of Taylor's *MA* scale without any buffer items were translated into Greek. They were then back translated into English for faithfulness of meaning to the original English statements. For those items in which there was a discrepancy between the back translation and the English statement, faithfulness of meaning was chosen over word-for-word translation. The Greek items were submitted to three judges (clinicians) who were instructed to sort them into two piles according to "comprehensibility" in Greek. The "noncomprehensible" items were rewritten in Greek in a simpler form. The rewritten items were again submitted to three judges who ranked them for faithfulness of meaning to the original Taylor items. For those items in which a discrepancy occurred, simplicity was chosen over faithfulness to translation.

The 50 items were then administered individually to a representative sample of 400 Athenians by specially trained interviewers. The interviewers were selected on the basis of intelligence tests, personality tests, and on their ability to establish an adequate rapport during surveys. They had been trained in survey techniques and had been interviewing biweekly for a minimum of 2 years. Continuous training and supervision were maintained. Fifteen percent of all interviews were checked by supervisors to insure maintenance of sampling procedure and quality of interviewing. Eleven interviewers were used for this study, aged 25–35, eight females and three males. The scale of social status developed by the Institute ranges from 1 to 9 (Triandis, Vassiliou, & Thomanek, 1966). The interviewers ranged from 3 to 5 on this scale.

SAMPLING PROCEDURE

The sample was chosen using a modified probability method developed and used before at the Athenian Institute of Anthropos. The population of Athens represents 23% of the population of Greece. The universe was adults (18 years of age and above) living in private households. Institutions such as hotels, hospitals, boarding schools, and military barracks were excluded. Persons sharing quarters and meals were considered to be members of households, and distinction was made between permanent members and temporary guests. Family links were not considered as a prerequisite for household participation. These definitions correspond to those used by the National Statistical Service.

According to the 1961 census the total population of the metropolitan area was found to be 1,850,709 people of whom 1,776,100 were found in private households (47% male, 53% female).

For the selection of the sampling unit the following steps were taken:

1. From a directory of all the streets of the Athens metropolitan area, using a table of random numbers, 50 streets were selected.

2. The selected streets were the "starting points" for each interviewer. Eight households were interviewed on each street. Four women and four men were selected from the first 15 starting points. Three males and five females were selected from the final 25 starting points. This procedure was followed to insure a 53–47% female-male representation.

3. On each street the interviewer followed at times the right and at times the left side of the street according to instructions based on alternate patterns.

4. The interviewer selected every third residential building. Stores, offices, etc., were excluded from the counting.

5. Where the selected building had more than one floor, the interviewer was instructed to follow a rotating pattern. According to this pattern, if the number of floors in selected consecutive houses increased, the selected floor was one higher than that previously selected. If the succeeding house had fewer floors than the previous one, the selected floor was again the first inhabited.

TABLE 1

MEAN *MA* SCALE SCORES OF THE REPRESENTATIVE SAMPLE

	M
Sex	
Males	17.14
Females	22.31
Age	
18–34	19.83
35–44	18.84
45 and over	20.53
Education[a]	
Illiterate	23.04
Grammar	20.54
High school	18.71
University	15.36
Income	
Less than 3000 drachmas	22.03
More than 3000 drachmas	18.50

[a] Illiterate is defined as up to 2 years of grammar school. Grammar includes up to 6 years of education. One or 2 years of grammar school education is classified under grammar. High school includes 12 years of education. University includes university education beyond high school.

6. If, on the selected floor, there was more than one apartment, the first apartment to the right of the staircase was selected.

7. The interviews were carried out from 10 A.M. to 6 P.M. (The Greek working day breaks from 2 P.M. to 5 P.M.)

8. The interviewer listed all the members of the household, oldest to youngest, according to their age. The youngest present who belonged to the sampling universe (18 years and above) was given the *MA* scale orally. The *MA* scale was part of a longer questionnaire involving general socioeconomic questions.

The proportions of sex, age, education, and income obtained in the sample corresponded to those obtained in over 100 public-opinion surveys conducted in the past 4 years. These proportions also corresponded to those of the National Statistical Service.

DATA ANALYSIS

The mean *MA* scale scores according to sex, three age levels, four educational levels, and two income groups were obtained. These results showed promising differences. In order to analyze any possible interaction effects, it was decided to further analyze the effects of sex, age, and education[1] using a three-way analysis of variance, fixed-constants model, with 10 different individuals per cell (McNemar, 1962). Because there were not enough subjects from the representative sample to fill each cell, particularly females-old-high education and both sexes-young-illiterate, additional subjects were interviewed. These additional subjects were selected

---

[1] Income was excluded from the analysis of variance. Income and education have been found to be significantly correlated in our population.

TABLE 2

ANALYSIS OF VARIANCE OF *MA* SCALE SCORES OF TEN SUBJECTS PER CELL FOR TWO SEX LEVELS, THREE AGE LEVELS, AND FOUR EDUCATIONAL LEVELS

Source	*Ss*	*df*	*MS*	*F*
Age (A)	160.03	2	80.02	1.20
Sex (S)	735.00	1	735.00	10.98*
Education (E)	1,954.70	3	651.60	9.73**
A × S	24.33	2	12.17	0.18
S × E	48.68	3	16.23	0.24
A × E	335.25	6	55.88	0.83
A × S × E	273.79	6	45.63	0.68
Within cells	14,462.20	216	66.95	
Total	17,993.98	239		

* $p < .01$.
** $p < .001$.

1 month after the initial interview using the same sampling procedure. Fifty additional starting points were selected which gave 400 new interviewees.[2]

RESULTS

Table 1 presents the mean *MA* scale scores of the representative sample. Note the higher *MA* scale scores in females compared to males. Age does not appear to affect *MA* scale scores. Note the decreasing *MA* scale scores with increased education. High income groups have lower *MA* scale scores than low income groups.

Table 2 indicates the results of the analysis of variance. Age was not a source of variance. Sex was a source of variance ($F = 10.98$, $p < .01$). Education was also a source of variance ($F = 9.73$, $p < .001$). The two-way and three-way interactions were not significant. This lack of significant interaction indicates that variation due to sex and education is consistent within age groups. Similarly sex varies independently of education and vice versa.

DISCUSSION

The results are clear-cut. Females have higher anxiety scores than males at all age levels and at all levels of education. The higher the educational level, the lower the anxiety score, at all age levels and for both sexes. This is illustrated in Figure 1.

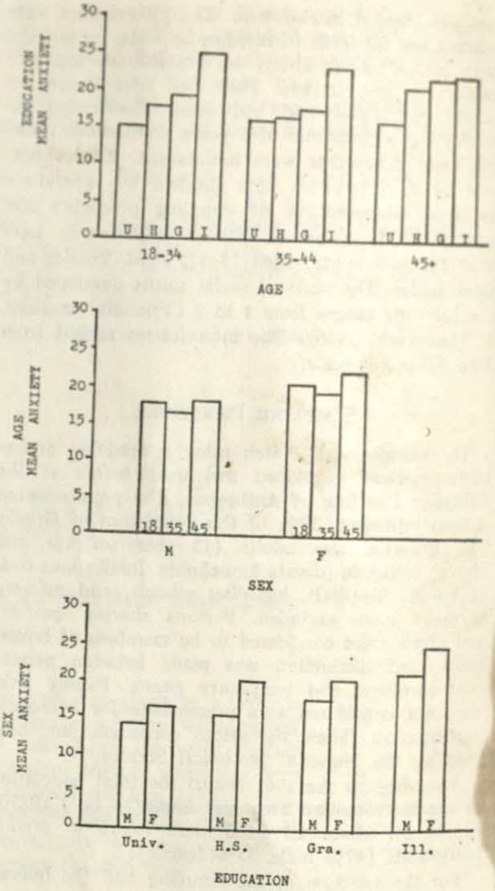


FIG. 1. Mean *MA* scale scores of Age × Education, Sex × Age, and Education × Sex from analysis of variance.

Whether the higher *MA* scale scores in women are due to biological or psychogenic factors or to the role culture assigns to women cannot be answered by this study. Comparison of Athenian and American subjects is not possible because in the relevant American studies in which sex differences were noted (Sarason et al., 1960); Sinick, 1956), the samples were nonrepresentative. A comparable sample of American subjects and possibly another culture might clarify the role of cultural determinants as sources of variation in sex.

The variation due to educational level is an important finding. It is unlikely that literacy affected the scores since the scale was administered orally. Why should high anxiety level be inversely related to high educational achievement? Is one possible answer that the educated person is more aware of the actual content of the scale resulting in an attempt to present a better façade? A partial answer might be provided by lie items which were not included in this form.[3]

---

[2] Since these surveys are conducted monthly, the starting points each month are selected to permit addition of successive samples. This procedure also permits trend and cumulative analysis (Vassiliou, Vasso, Vlachos, and Vassiliou, "On Leisure Time," in preparation).

Education as a source of variation must be viewed within the total cultural context. Holding education constant and varying the cultural setting from a relatively highly developed urban area to a rural setting would provide information here.

If the *MA* scale measures "drive level," why should subjects of high educational achievement have a lower anxiety level than subjects of low educational achievement? Raphelson (1957) found that a measure of achievement correlated −.43 and −.25 with the Mandler-Sarason Test Anxiety Scale and the *MA* scale, respectively. Kausler and Trapp (1958) also found a negative relationship between the *MA* scale and the French Achievement Motivation Test. These negative relationships are consistent with the results obtained in this study. These results suggest that high drive level, as measured by the *MA* scale, is related to low educational achievement. Holding education constant would clarify the relationship between drive level and achievement.

Interviewer bias might be taken into account as a possible source of variation in *MA* scale scores. The interviewers were selected to be a homogeneous group ranging from 3 to 5 on a 9-point scale of social status. Fifty-five percent of the population falls within the 3–5 point range, indicating the interviewers represented the central and majority sections of the population. If there was an interviewer bias, it affected the lower and higher levels of social status. Because the interviewers were such a homogeneous group, interviewer social status as a source of variation could not be tested. This would require a design with interviewers of disparate social status.

Interpretation of the results must also account for the fact that the *MA* scale was orally administered. A paper-and-pencil form of administration may yield different results. Research in progress has suggested this possibility.

The finding that anxiety scores vary according to sex and education has implications for personality research. Studies of neurotics or psychotics which do not account for the effect of sex or education on anxiety level can be misleading. If, for example, we were to compare normal college students with mental patients from a relatively low educational level, differences in anxiety level would be found almost certainly. But this difference in anxiety level would not necessarily be due to personality integration.

REFERENCES

KAUSLER, D. H., & TRAPP, E. P. Relationship between achievement motivation and manifest anxiety scores. *Journal of Consulting Psychology,* 1958, **22**, 448–450.

MCNEMAR, Q. *Psychological statistics.* (3rd ed.) New York: Wiley, 1962.

RAPHELSON, A. C. The relationships among imaginative, direct verbal, and physiological measures of anxiety in an achievement situation. *Journal of Abnormal and Social Psychology,* 1957, **54**, 13–18.

SARASON, S. B., ET AL. *Anxiety in elementary school children: A report of research.* New York: Wiley, 1960.

SINICK, D. Two anxiety scales correlated and examined for sex differences. *Journal of Clinical Psychology,* 1956, **12**, 394–395.

TAYLOR, J. A. A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology,* 1953, **48**, 285–290.

TRIANDIS, H. C., VASSILIOU, V., & THOMANEK, E. K. Social status as a determinant of respect and friendship acceptance. *Sociometry,* 1966, **29**, 396–405.

(Received May 12, 1966)

---

[3] Lie items are being included in the new revision which has been given to representative samples of other urban and rural areas in Greece.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 198-202

VERBAL OPERANT CONDITIONING, INTERTRIAL ACTIVITY, AWARENESS, AND THE EXTENDED INTERVIEW

WILLIAM F. OAKES

University of Hawaii

Four groups of 19 Ss constructed sentences using a verb and 1 of 6 pronouns presented on an index card in a series of 6 blocks of 20 trials. In a 2 × 2 design, 2 of the groups were reinforced for sentences begun with "I" or "we," while the other two were not reinforced, and 2 of the groups were required to name colors between trials, while 2 were not. After the conditioning trials, Ss were separated into "aware" and "unaware" Ss on the basis of the Spielberger (1962) interview and criteria. It was found that: (a) The reinforced Ss gave more "I-we" sentences than the unreinforced; (b) the aware Ss gave more "I-we" sentences than the unreinforced control Ss; (c) the unaware Ss gave more "I-we" sentences than the unreinforced control Ss; and (d) the aware and unaware Ss did not differ in the number of "I-we" sentences given. It was concluded that although findings a and b are consistent with either the cognitive or the reinforcement interpretation of verbal operant conditioning, findings c and d are consistent with the reinforcement interpretation and contrary to the cognitive interpretation. Thus the results suggest that the effect of reinforcement on the response class in verbal operant conditioning under certain conditions is direct, i.e., not mediated by cognitive processes.

Dixon and Oakes (1965) have shown that in the verbal operant conditioning (VOC) sentence-construction task, if the opportunity of a subject to formulate hypotheses during acquisition regarding the response-reinforcement contingency is interfered with, the usual relationship between awareness and VOC is not found, although conditioning is still evident. For half their subjects, Dixon and Oakes required a color-naming task between trials. Using the Dulany (1962) concepts of Reinforcement Hypothesis (RH), Behavioral Hypothesis (BH), and Behavioral Intention (BI) as indexes of "awareness," the typical correlation between VOC and awareness of the response-reinforcement contingency was found for the subjects who did not have the color-naming task between trials, while a nonsignificant correlation in the opposite direction (significantly different from that for the non-color-naming subjects) was found for the color-naming subjects. There was no difference in the amount of conditioning of the subjects with and without color naming.

The results of the Dixon and Oakes experiment were interpreted as being incompatible with the position of Dulany (1962) and Spielberger (1962), who held that the subject's increase in emission of members of the reinforced-response class in VOC was *necessarily* mediated by cognitive processes, that is, awareness by the subject of the response-reinforcer contingency, and a conscious intention to comply with the experimenter's perceived wishes. Rather, the results were interpreted as being consistent with the position of Verplanck (1962) and Krasner (1962), who held that there is no *necessary* relationship between awareness and VOC, that is, that the effect of the reinforcer is direct, and not necessarily mediated by cognitive processes.

Spielberger (1962) and his students (e.g., DeNike, 1964; Levin, 1961) have argued that the usual techniques used for assessing "awareness" in VOC experiments, that is, a short interview of four or five questions, is inadequate for identifying all the subjects who are *really* aware, and have used an extended interview (EI) of some 16 increasingly pointed questions to separate their "aware" and "unaware" subjects. The Dulany concepts of RH, BH, and BI are each assessed by four or five questions, that is, by short interviews. The RH concept, which gets at the subject's perception or awareness of the response-reinforcement contingency, is the only one of the Dulany concepts which assesses what we usually have in mind when we think of "awareness" in the context of VOC. His BH has to do with the subject's interpretation of the experimenter's motivation, that is, that the experimenter wants the subject to use the words reinforced, and his BI has to do with the subject's intention to comply. Thus the BH and BI concepts are getting at what Orne (1962) referred to as the "demand characteristics" of the experiment, and not at awareness per se.

Since RH, the degree of awareness in the Dulany scheme, is assessed with only a five-question interview, the Spielberger criticism could be made, that one cannot really separate the aware and unaware subjects on that basis. Of course, by extending the Spielberger argument one could maintain that you cannot *really* identify *all* the truly aware subjects without using an interview of infinite length. And since this is absurd, the approach of attempting to correlate degree of awareness with degree of VOC is likely to be more productive than attempting to identify aware and unaware subjects.

The Dixon and Oakes study demonstrated not that VOC occurred without awareness, but rather that the relationship between degree of VOC and degree of awareness (RH) is a function of whether or not the subject is permitted time to formulate and test hypotheses between trials during acquisition. There is the implication, however, that since there was no correlation between RH and VOC for the color-naming subjects, such learning may occur without awareness. The present experiment is an attempt to get at that implication using the more sensitive Spielberger EI to separate the aware and unaware subjects. We recognize, of course, that no such experiment can provide completely definitive results, since an interview of infinite length is not used. We can, however, answer the question of whether interfering with the subject's opportunity to formulate and test hypotheses during acquisition produces results consistent with those that Spielberger and his students have obtained without such interference.

The design of this experiment thus involves the sentence-construction VOC paradigm, with half the subjects having the intertrial activity (color naming), and with the Spielberger EI used afterward to separate the aware and unaware subjects.

METHOD

Subjects. The subjects were 76 students from introductory psychology classes at the University of Hawaii. They were randomly assigned to four groups of 19 each: two experimental groups which both received reinforcement, one group with color naming between trials, the other without, and two control groups, neither of which was reinforced, one with the color-naming task between trials, the other without.

Apparatus. The stimulus materials for the conditioning task consisted of 3 × 5-inch white index cards, with a different neutral past-tense verb typed on each. These verbs were taken from the list used by Binder and Salop (1961). The pronouns I, we, you, he, she, and they were typed above the verb with the order of appearance of the pronouns randomized over all cards. The subject was presented with easily recognizable colors shown in the aperture of a memory drum. The colors were arranged in random order by selection from a large assortment of colors.

A screen with its leaves folded in at approximately 120 degrees toward the subject was used to separate the subject from the experimenter. The central portion of the screen was 26 × 30 inches, and the leaves were 12 × 30 inches. A place for the presentation of the card was made 12 inches above the base of the screen and in the center of the central panel. Beneath the hole through which the card was presented, a slot was cut in the screen to fit the memory drum so that the memory drum would face the subject. The colors on the memory drum thus appeared directly below the card holder.

The experimenter used a scoring sheet on which the numbers 1–120 were followed by the pronouns used in the experiment, and he circled the single pronoun used by the subject on each trial on that data sheet.

Procedure. Subjects were seen individually by the experimenter.[1] As the subject arrived at the experimental room the experimenter engaged the subject in a short informal conversation in order to establish a friendly relationship and to put the subject at ease. After this informal interview the usual instructions for the Taffel (1955) procedure were read to the subject. Subjects in the color-naming groups were instructed to name the colors appearing on the memory drum between trials.

The subjects then proceeded to construct sentences from the 120 cards on which the pronouns and the verb had been typed. For the reinforced groups, reinforced control (RC) and reinforced naming (RN), during the first 20 trials every third sentence constructed by the subject was followed by the experimenter mumuring "mm-hmm," regardless of which pronoun was used. This was done so that the first reinforcement after the first block of 20 trials, the operant-level trials, would not be such a surprise to the subject (Matarazzo, Saslow, & Pareis, 1960). After the first 20 trials, each time the subject began a sentence with either "I" or "we," the experimenter said "good" at the end of the sentence for both reinforced groups, RC and RN.

For the nonreinforced groups, unreinforced control (UC) and unreinforced naming (UN), the experimenter remained quiet throughout the 120 trials.

The color-naming task performed by the color-naming groups, RN and UN, consisted of naming colors during the 10-second intertrial period. The colors appeared in the memory drum every 1.59 seconds. The non-color-naming groups, RC and UC, did not perform any task during the 10-second intertrial interval.

---

[1] The writer would like to express his appreciation to Falak Thaver, who served as the experimenter in this study.

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF NUMBER OF "I-WE" SENTENCES OF ALL SUBJECTS ON THREE VARIABLES: A, REINFORCEMENT; B, INTERTRIAL ACTIVITY; AND C, BLOCKS OF TRIALS

Source	*df*	*MS*	*F*
Between *Ss*			
A	1	1195.46	26.26**
B	1	177.89	3.91
A × B	1	72.52	1.59
Error (between)	72	45.52	
Within *Ss*			
C	4	6.10	
A × C	4	22.58	3.62**
B × C	4	2.67	
A × B × C	4	5.00	
Error (within)	288	6.23	

** $p < .01$.

Following the presentation of the 120 cards all subjects received the postconditioning interview, which was a series of questions read to the subject by the experimenter one at a time, with the answers to the questions being written by the subject on a blank sheet of paper. The interview used was the one used by Spielberger (1962).[2]

Two graduate students in psychology were used as judges to rate the interview protocols. Both judges had previously conducted VOC research and were familiar with the literature in the area. Interview protocols of all reinforced subjects were given the two judges for independent ratings. The Spielberger criteria for aware and unaware subjects were used, and agreement by the two judges was required for inclusion of the cases in analyses involving awareness.

RESULTS

The 120 trials were divided into six blocks of 20 trials each. The number of "I-we" sentences was taken as the dependent variable. The score on the first (unreinforced) block of trials was considered the operant level for each subject. An analysis of variance of operant levels indicated that neither the main effects nor the interaction approached significance, all *F* values being less than 1.00. All further analyses were made using scores on the second through sixth (i.e., the reinforced) blocks of trials.

Using all subjects, an analysis of variance was made involving three variables: reinforcement, intertrial activity, and blocks of trials. As can be seen from the summary of that analysis in Table 1, there was a significant reinforcement effect, $F\ (1,\ 72) = 26.26,\ p < .01$, and interaction of reinforcement and blocks of trials, $F\ (4,\ 288) = 3.62,\ p < .01$. Thus the reinforced subjects gave more "I-we" sentences, increasing over blocks of trials, while the unreinforced subjects tended to decrease over blocks of trials. No other effects were significant.

Regarding awareness, the judges agreed in their judgments of 6 subjects from Group RN as aware and 11 as unaware, and 7 subjects from Group RC as aware and 10 as unaware. Since the experimenter remained silent even on the operant trials for the unreinforced subjects (Groups UN and UC), and there was thus no contingent stimulus, no subjects in those groups were considered aware.

To answer the question of whether the aware subjects conditioned, an unweighted-means analysis of variance was made involving the subjects judged aware and their unreinforced controls, Groups UN and UC. This analysis involved three variables: reinforcement, intertrial activity, and blocks of trials. The summary of that analysis in Table 2 indicates that there were significant main effects of reinforcement, $F\ (1,\ 47) = 23.46,\ p < .01$, intertrial activity, $F\ (1, 47) = 4.63,\ p < .05$, and blocks of trials, $F\ (4,\ 188) = 2.41,\ p < .05$. Also significant was the interaction of reinforcement and blocks of trials, $F\ (4,\ 188) = 4.89,\ p < .01$. Thus, the aware subjects gave more "I-we" sentences than their unreinforced controls overall, subjects with the

TABLE 2

SUMMARY OF UNWEIGHTED-MEANS ANALYSIS OF VARIANCE OF NUMBER OF "I-WE" SENTENCES OF AWARE SUBJECTS AND UNREINFORCED CONTROL SUBJECTS ON THREE VARIABLES: A, REINFORCEMENT; B, INTERTRIAL ACTIVITY; AND C, BLOCKS OF TRIALS

Source	*df*	*MS*	*F*
Between *Ss*			
A	1	788.65	23.46**
B	1	155.65	4.63*
A × B	1	81.58	2.43
Error (between)	47	33.61	
Within *Ss*			
C	4	14.44	3.41*
A × C	4	29.23	4.89**
B × C	4	8.85	1.48
A × B × C	4	8.89	1.49
Error (within)	188	5.98	

* $p < .05$.
** $p < .01$.

---

[2] Any possible experimenter bias would have to enter at this point, since the experimenter did know the hypothesis being tested. However, it is believed that this is unlikely, as the same interview schedule was used for all subjects with written responses by the subject, and awareness was judged by different individuals who judged the protocols blind.

TABLE 3

SUMMARY OF UNWEIGHTED-MEANS ANALYSIS OF VARIANCE OF NUMBER OF "I-WE" SENTENCES OF UNAWARE SUBJECTS AND UNREINFORCED CONTROL SUBJECTS ON THREE VARIABLES: A, REINFORCEMENT; B, INTERTRIAL ACTIVITY; AND C, BLOCKS OF TRIALS

Source	*df*	*MS*	*F*
Between *S*s			
A	1	510.48	15.17**
B	1	96.79	2.88
A × B	1	33.20	
Error (between)	55	33.66	
Within *S*s			
C	4	3.71	
A × C	4	4.04	
B × C	4	0.41	
A × B × C	4	3.01	
Error (within)	220	5.19	

** $p < .01$.

intertrial activity gave fewer "I-we" sentences overall than those without, and there was an overall increase in "I-we" sentences over blocks of trials. The significant Reinforcement × Blocks of Trials interaction indicates an increase over blocks of trials for reinforced subjects and a tendency for unreinforced subjects to decrease over blocks of trials.

To determine whether the unaware subjects conditioned, a similar unweighted-means analysis of variance was made involving the subjects judged unaware and their unreinforced controls. Table 3 presents the results of this analysis. It can be seen that the only significant effect was reinforcement, $F\ (1,\ 55) = 15.17,\ p < .01$. The significant reinforcement effect is the result of the unaware reinforced subjects giving more "I-we" sentences than their unreinforced controls, as can be seen from Table 5.

TABLE 4

SUMMARY OF UNWEIGHTED-MEANS ANALYSIS OF VARIANCE OF NUMBER OF "I-WE" SENTENCES OF AWARE AND UNAWARE REINFORCED SUBJECTS ON THREE VARIABLES: A, INTERTRIAL ACTIVITY; B, AWARENESS; AND C, BLOCKS OF TRIALS

Source	*df*	*MS*	*F*
Between *S*s			
A	1	249.29	4.92*
B	1	67.10	1.32
A × B	1	14.38	
Error (between)	30	50.70	
Within *S*s			
C	4	20.83	2.23
A × C	4	8.03	
B × C	4	16.99	1.82
A × B × C	4	7.74	
Error (within)	120	9.36	

* $p < .05$.

Table 4 presents the results of an analysis of variance involving only the reinforced subjects, with the variables of intertrial activity, awareness, and blocks of trials. It can be seen that the only significant effect is that of intertrial activity, $F\ (1,\ 30) = 4.92,\ p < .05$, indicating a tendency for subjects with intertrial activity to give fewer "I-we" sentences overall than those without.

Table 5 presents the cell means involved in all comparisons in the above analyses, permitting inspection of the directions of the differences found significant in the analyses of variance.

DISCUSSION

The results obtained clearly indicate (*a*) that the reinforcement was effective, that is, the reinforced subjects gave more "I-we" sentences than the unreinforced subjects, (*b*) that the aware

TABLE 5

MEAN NUMBER OF "I-WE" SENTENCES FOR ALL GROUPS ON SIX BLOCKS OF TRIALS

Groups[a]	Blocks					
	1	2	3	4	5	6
RN ($N = 19$)	6.05	6.94	7.31	8.00	9.57	8.31
A ($N = 6$)	5.67	7.00	7.00	6.66	10.33	9.50
U ($N = 11$)	6.45	6.73	7.64	7.64	8.36	6.64
RC ($N = 19$)	5.89	8.89	10.36	10.73	10.52	10.84
A ($N = 7$)	5.00	8.14	11.14	12.28	11.71	12.71
U ($N = 10$)	6.20	9.20	9.70	9.30	9.30	9.00
UN ($N = 19$)	5.31	5.84	5.58	5.10	4.94	5.31
UC ($N = 19$)	5.89	6.10	6.21	5.68	5.89	5.37

Note.—Trial Block 1 provided the operant level for reinforced groups (RN and RC), with "I-we" sentences reinforced for those subjects in blocks 2 through 6.

[a] RN: Reinforced, Naming; A: Aware; U: Unaware; RC: Reinforced, Control; UN: Unreinforced, Naming; UC: Unreinforced, Control.

subjects gave more "I-we" sentences than the unreinforced control subjects, (*c*) that the unaware subjects gave more "I-we" sentences than the unreinforced control subjects, and (*d*) that the aware and unaware subjects did not differ in the number of "I-we" sentences given. Findings *a* and *b* are consistent with either the cognitive (i.e., Dulany, 1962; Spielberger, 1962) interpretation or the reinforcement (i.e., Krasner, 1962; Verplanck, 1962) interpretation of VOC. However, findings *c* and *d* are clearly contradictory to the cognitive interpretation and consistent with the reinforcement interpretation.

The present findings are consistent with the results of the Dixon and Oakes (1965) study in all essential details. We thus are led to the conclusion that VOC is not *necessarily* mediated by cognitive processes (i.e., awareness), but rather that under certain conditions (e.g., interference with the subject's opportunity to formulate and test hypotheses during acquisition) the effect of the reinforcer may be direct. As Dixon and Oakes (1965) suggested,

> When conditions are such that *S* is not able to formulate hypotheses during acquisition, the same degree of learning may be exhibited as when *S* does have the opportunity to formulate hypotheses, indicating a direct strengthening effect of the reinforcer, which is not correlated with the degree of *S*'s awareness. However, when such *S*s are later interviewed, as a result of the nature of the experimental task and the interview, a certain distribution of "awareness" levels will be exhibited by *S*s. But this "awareness" on the part of individual *S*s, which may indeed be produced by the interview itself, need not be related to the degree of conditioning exhibited by the particular *S*s during the experiment [p. 157].

REFERENCES

BINDER, A., & SALOP, P. Reinforcement and personality factors in verbal conditioning. *Journal of Psychology,* 1961, 52, 379–402.

DENIKE, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology,* 1964, 68, 521–529.

DIXON, P. W., & OAKES, W. F. Effect of intertrial activity on the relationship between awareness and verbal operant conditioning. *Journal of Experimental Psychology,* 1965, 69, 152–157.

DULANY, D. E., JR. The place of hypotheses and intentions: Analysis of verbal control in verbal conditioning. In C. W. Ericksen (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962.

KRASNER, L. Reported awareness in verbal conditioning as a function of experimental conditions and subjects' personality. Paper read at Western Psychological Association, San Francisco, April, 1962.

LEVIN, S. M. The effects of awareness on verbal conditioning. *Journal of Experimental Psychology,* 1961, 61, 67–75.

MATARAZZO, J. D., SASLOW, J., & PAREIS, E. N. Verbal conditioning of two response classes: Some methodological considerations. *Journal of Abnormal and Social Psychology,* 1960, 61, 190–206.

ORNE, M. T. On the social psychology of the psychological experiment with particular reference to demand characteristics and their implications. *American Psychologist,* 1962, 17, 776–783.

SPIELBERGER, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology,* 1955, 51, 496–501.

VERPLANCK, W. S. Unaware of where's awareness: Some verbal operants—notates, monents, and notants. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962.

(Received May 23, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 203–206

VERBAL CONDITIONING IN SOCIOPATHS WITH SOCIAL AND MONETARY REINFORCEMENT

J. L. BERNARD AND RUSSELL EISENMAN
Milledgeville State Hospital, Georgia *Temple University*

Since it has been maintained that sociopaths have a learning deficit, an attempt was made to condition sociopaths and normal Ss verbally with "good" (social reward) or nickels (monetary reward). 40 female prisoners comprised the sociopathic group and were chosen from a larger group of inmates because of their sociopathic MMPI profiles. 39 normal Ss were chosen from a larger group of student nurses; all normal Ss had MMPI T scores between 35 and 65, without sociopathic or psychotic tendencies on the MMPI. Sociopaths showed significantly more frequent emission of the reinforced pronoun in the sentence-construction task. Social reinforcement was more effective than monetary rewards for both groups, but noticeably so for the sociopaths. Possible bases of the results are discussed.

The term "psychopath" or sociopath" may represent little more than a wastebasket category into which diverse personality types are placed (Pennington, 1954; White, 1956). Nevertheless, definite opinions have been formed about the sociopathic personality. Bender (1947) considered sociopaths unsuitable candidates for psychotherapy because of their inability to enter into relationships or profit from education. Consistent with Bender's view is Cleckley's (1950) concept of semantic dementia, which suggests that sociopaths fail to react to words which bring about emotional reactions in normal people. For example, the sociopath may intellectually comprehend the statement "I'm ashamed of you," but he lacks any emotional appreciation of its meaning. Evidence (Johns & Quay, 1962; Lykken, 1957; Quay & Hunt, 1965) has been presented which supports Cleckley's view that sociopaths have some form of learning deficit: "The psychopath can thus be said to be one who knows the words but not the music [Johns & Quay, 1962, p. 217]."

Despite the above, the psychopath or sociopath remains something of an enigma, in part because some of the evidence (Johns & Quay, 1962; Lykken, 1957; Quay & Hunt, 1965) has been questioned on methodological grounds (Persons & Persons, 1965)[1] and in part due to reports of psychotherapeutic success with sociopathic patients (Caditz, 1961; LaBarba, 1965; Persons, 1965; Schmideberg, 1947; Shelly & Johnson, 1961; Sturup, 1952).

The present research represents an attempt to condition sociopaths verbally with two kinds of reinforcement: social and monetary. When we say "good" to some individual we are employing a social reinforcement, and the behavior which we attempt to condition will probably have its primary usefulness in a social context (London & Rosenhan, 1964). Thus, if sociopaths are deficient in social learning, it may be that they would learn for something other than the personal social reward. Consequently, the present study used both the personal social reinforcer, "good," and the impersonal monetary reinforcer, nickels, in order to see if sociopaths could be conditioned.

METHOD

Subjects

The MMPI was used to select sociopathic subjects from a group of 85 female prisoners at the Georgia Woman's Prison. The MMPIs of all 85 were checked, and 40 with a sociopathic profile were chosen. Potential subjects were discarded if they showed psychotic profiles or if they failed to have 4–9 and 9–4 (sociopathic) profiles (Hathaway & Meehl, 1951). Thus, the final 40 sociopathic subjects had behavioral indicants of sociopathy which brought them to prison in the first place and also had sociopathic profiles on the MMPI, without psychotic tendencies as revealed by that test.

The normal group was selected from a class of 90 females. These subjects were student nurses who were administered the MMPI as part of the present research. From the 90 nurses, 39 were obtained whose MMPI scores were all between T scores of 35 and 65, without psychotic or sociopathic profiles.

Procedure

Both groups were administered a verbal conditioning task which was a variant of Taffel's (1955)

---

[1] Criticism of the Quay and Hunt (1965) research is based on an extended report, which is open to the same criticism that Persons and Persons (1965) made against the Johns and Quay (1962) study, according to a personal communication from Roy Persons, November 15, 1965.

TABLE 1

MEAN NUMBER OF "I" CHOICES IN BLOCKS OF TEN TRIALS FOR SOCIOPATHS AND NORMALS

	Condition 1[a]			Condition 2[b]			
Blocks	1–2	3–8	9–10	1–2	3–5	6–8	9–10
Sociopaths							
M	3.72	6.31	6.28	3.59	5.66	6.33	6.05
SD	1.61	0.76	0.84	0.92	0.81	1.03	0.98
Normals							
M	2.72	3.62	3.33	2.85	4.33	3.50	3.68
SD	0.92	0.71	0.79	0.68	0.65	0.59	0.53
	Condition 3[c]			Condition 4[d]			
Blocks	1–2	3–8	9–10	1–2	3–5	6–8	9–10
Sociopaths							
M	4.22	5.01	5.50	3.86	5.33	6.63	6.59
SD	0.44	0.88	0.70	0.57	1.00	1.32	1.13
Normals							
M	2.90	2.85	2.40	3.65	3.27	3.85	4.00
SD	0.39	0.63	0.57	0.58	0.81	0.72	0.81

[a] Operant level blocks 1–2, social reinforcement blocks 3–8, extinction blocks 9–10.
[b] Operant level blocks 1–2, social reinforcement blocks 3–5, monetary reinforcement blocks 6–8, extinction blocks 9–10.
[c] Operant level blocks 1–2, monetary reinforcement blocks, 3–8, extinction blocks 9–10.
[d] Operant level blocks 1–2, monetary reinforcement blocks 3–5, social reinforcement blocks 6–8, extinction blocks 9–10.

technique. The subjects were presented with 100 sentences typed on 3 × 5-inch cards, one sentence per card. On all sentences the first word was missing, and the subject was instructed to begin the sentences with I, we, he, she, or they. These pronouns were written on separate cards which were kept before the subject throughout the experiment.

All subjects were given 100 trials. The first 20 trials were used to establish an operant level for the pronoun ("I") which was to be reinforced, trials 21–80 were the reinforcement trials, and trials 81–100 were extinction trials in which no reinforcement was provided. On trials 21–80 one of two types of reinforcement was provided: social reinforcement in which the experimenter said "good" if the subject began a sentence with the pronoun "I," and monetary reinforcement in which the experimenter dropped a nickel into a small bowl, placed beside the subject, whenever the subject used "I" to begin a sentence. To study the effects of shifting from one kind of reinforcement to another, there were four conditions:

Condition 1. Trials 21–80, social reinforcement "good." Normal $N = 9$, sociopathic $N = 9$.

Condition 2. Trials 21–50, social reinforcement "good." Trials 51–80, nickels. Normal $N = 10$, sociopathic $N = 11$.

Condition 3. Trials 21–80, nickels. Normal $N = 10$, sociopathic $N = 9$.

Condition 4. Trials 21–50, nickels. Trials 51-80, social reinforcement "good." Normal $N = 10$, sociopathic $N = 11$.

The subjects were tested by five male experimenters, with the assignment of subjects to a given experimenter being random. The experimenters never informed the subjects that they could keep the money, and at the end of the experiment the nickles were retained, and the purpose of the study briefly explained to protesting subjects. At the conclusion of the experiment for a given subject, she was ushered out past waiting subjects. Thus, subjects who completed the experimental procedure could not communicate procedure to subjects who were waiting to participate.

RESULTS

Table 1 shows the mean number of "I" choices for the four conditions. The results were analyzed by a 2 (Groups) × 4 (Conditions) analysis of variance, which is summarized in Table 2. Sociopaths conditioned more than normals ($F = 18.20$, $df = 1/64$, $p < .01$). The significant A × B interaction indicated that sociopaths showed a more frequent emission of the word "I" under social as opposed to monetary reinforcement ($F = 31.00$, $df = 3/64$, $p < .01$).

TABLE 2

ANALYSIS OF VARIANCE FOR VERBAL CONDITIONING OF SOCIOPATHS AND NORMALS IN FOUR CONDITIONS

Source	*df*	*MS*	*F*
Groups (A)	1	91	18.20*
Conditions (B)	3	10	2.00
A × B	3	155	31.00*
Error	64	5	

* $p < .01$.

Mann-Whitney U tests for the four conditions all equaled 0 ($p < .001$ for all conditions), indicating that there was no overlap. In other words, sociopaths showed more frequent emission of "I" responses under all conditions with social or monetary reinforcement; as the analysis of variance revealed, social reinforcement was more effective than monetary reinforcement in conditioning sociopaths.

After the experimental conditions had been imposed, subjects were interviewed briefly to ascertain if they knew what the experimenter was doing or why he acted as he did. With this brief interview, no subjects were able to report the reinforcement contingency, although more extensive questioning as suggested by Levin (1961) may have shown signs of "awareness."

DISCUSSION

Sociopathic subjects conditioned better than normal subjects for either social or monetary reinforcement. While some might predict that sociopaths would condition for monetary rewards, the sociopaths' conditioning to the more personal social reinforcement ("good") contradicts what many would expect. It would appear that the sociopathic female prisoners were not impaired with any semantic dementia and were able to respond to reinforcement even when it was social in nature (London & Rosenhan, 1964). The sociopaths' more frequent use of "I" throughout the experiment is actually similar to the results found by Johns and Quay (1962), although their study has been interpreted as support for the view that sociopaths failed to condition. As Persons and Persons (1965) have pointed out, the sociopaths in the Johns and Quay study were superior in performance throughout the experiment, although their change from their initially high level was not as great as the initially low-scoring neurotics.

The superiority of the social reinforcer "good" over the nickels may have been due to the high value that praise has in our society, in which case the sociopaths may have been especially influenced by praise from opposite-sex experimenters, due to confinement making opposite-sex contact impossible or limited only to visitors. It is also possible that the result of not informing subjects that they would be given the nickels was that they did not think of the money as reinforcement.

If the present study casts doubt on the validity of Cleckley's (1950) semantic dementia concept for explaining the behavior of sociopaths, what other explanations can account for their frequent antisocial behavior? Perhaps what occurs is that sociopaths have learned different values, imitate different models, and otherwise learn a different pattern of behavior than the normal population. Social rewards may or may not be effective with any given sociopathic population, depending upon the value of the particular social reward administered. As several investigators have suggested, verbal conditioning can be affected by situational variables (Ebner, 1965; Kanfer & Karas, 1959; Weiss, Krasner, & Ullmann, 1960). Thus, it is probably a mistake to assume that sociopaths will necessarily condition or not condition. The present study revealed that sociopaths were superior to normal subjects in verbal conditioning, but situational or subject variables may account for the differences found here from findings of Johns and Quay (1962) and Quay and Hunt (1965). Thus, while it seems established that sociopaths fail to learn in a variety of situations (Hare, 1965a, 1965b, 1965c; Lykken, 1957; McCord & McCord, 1956), the mistake should not be made of considering sociopaths as necessarily unable to learn, profit from psychotherapy, or otherwise respond to social stimuli.

REFERENCES

BENDER, L. Psychopathic behavior disorders in children. In R. M. Linder & R. V. Seliger (Eds.), *Handbook of correctional psychology*. New York: Philosophical Library, 1947. Pp. 360–377.

CADITZ, S. B. Effects of a forestry camp experience on the personality of delinquent boys. *Journal of Clinical Psychology*, 1961, **17**, 78–81.

CLECKLEY, H. *The mask of sanity*. St. Louis: Mosby, 1950.

EBNER, E. Verbal conditioning in schizophrenia as a function of degree of social interaction. *Journal of Personality and Social Psychology*, 1965, **1**, 528–532.

HARE, R. D. Acquisition and generalization of a conditioned-fear response in psychopathic and nonpsychopathic criminals. *Journal of Psychology*, 1965, **59**, 367–370. (a)

HARE, R. D. Psychopathy, fear arousal, and anticipated pain. *Psychological Reports*, 1965, **16**, 499–502. (b)

HARE, R. D. Temporal gradient of fear arousal in psychopaths. *Journal of Abnormal Psychology*, 1965, **70**, 442–445. (c)

HATHOWAY, S. R., & MEEHL, P. E. *An atlas for the clinical use of the MMPI*. Minneapolis: University of Minnesota Press, 1951.

JOHNS, J. H., & QUAY, H. C. The effect of social reward on verbal conditioning in psychopathic and neurotic military offenders. *Journal of Consulting Psychology*, 1962, **26**, 217–220.

KANFER, F. H., & KARAS, S. C. Prior experimenter-subject interaction and verbal conditioning. *Psychological Reports*, 1959, **5**, 345–353.

LaBarba, R. C. The psychopath and anxiety: A reformulation. *Journal of Individual Psychology*, 1965, **21**, 167–170.

Levin, S. M. The effects of awareness on verbal conditioning. *Journal of Experimental Psychology*, 1961, **61**, 67–75.

London, P., & Rosenhan, D. Personality dynamics. *Annual Review of Psychology*, 1964, **15**, 447–492.

Lykken, D. F. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, 1957, **55**, 6–10.

McCord, W., & McCord, J. *Psychopathy and delinquency*. New York: Grune & Stratton, 1956.

Pennington, L. A. Psychopathic and criminal behavior. In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology*. New York: Ronald Press, 1954. Pp. 421–447.

Persons, R. W. Psychotherapy with sociopathic offenders: An empirical evaluation. *Journal of Clinical Psychology*, 1965, **21**, 204–207.

Persons, R. W., & Persons, C. E. Some experimental support of psychopathic theory: A critique. *Psychological Reports*, 1965, **16**, 745–749.

Quay, H. C., & Hunt, W. A. Psychopathy, neuroticism, and verbal conditioning. *Journal of Consulting Psychology*, 1965, **29**, 283.

Schmideberg, M. The treatment of psychopaths and borderline patients. *American Journal of Psychotherapy*, 1947, **1**, 45–70.

Shelly, E., & Johnson, W. F., Jr. Evaluating an organized counseling service for youthful offenders. *Journal of Counseling Psychology*, 1961, **8**, 351–354.

Sturup, G. K. The treatment of criminal psychopaths in Hertedvester. *British Journal of Medical Psychology*, 1952, **25**, 31–38.

Taffel, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology*, 1955, **51**, 496–501.

Weiss, R. L., Krasner, L., & Ullmann, L. P. Responsivity to verbal conditioning as a function of emotional atmosphere and pattern of reinforcement. *Psychological Reports*, 1960, **6**, 415–426.

White, R. W. *The abnormal personality*. New York: Ronald Press, 1956.

(Received May 26, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 206–211

DEVELOPMENT OF CONTRACTUAL NORMS IN A DYAD[1]

PETER MURDOCH[2]
University of North Carolina at Chapel Hill

This study investigated some circumstances leading to the development of norms which reduce threats to the continued existence of a profitable dyadic relationship. Pairs of Ss played a bargaining game in which 1 person was given the power to divide the joint rewards of the pair, while the other person had the power to leave the relationship for an attractive alternative. Pairs were given opportunities to form contractual norms specifying rules of fairness and loyalty. Perceptions of the unfairness of the reward divider and the disloyalty of the other member, as well as the precision with which the reward divider could allocate joint outcomes, were varied in a 2 × 2 × 2 factorial design. 12 dyads were run in each of the 8 conditions. The major predictions were confirmed. When the reward divider had precise control over the division of joint rewards, an Unfairness × Disloyalty interaction was found. Dyads formed the greatest number of contractual norms when the reward divider was presented as likely to exploit and when the other member was presented as likely to withdraw from the relationship. Where the reward divider had only imprecise control over the division of joint outcomes, the frequency of contracts was generally higher than in the precise control conditions.

This paper presents a brief theoretical and experimental analysis of how interdependent groups may respond to threats to their existence by developing norms which regulate the behavior of group members. For this context, norms

---

[1] This report is based on parts of a dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the PhD degree. This study was supported by the Organization Research Group operating under contract NONR-855(04) from the Office of Naval Research. The author wishes to give heartfelt thanks to John Thibaut, his dissertation director. Thanks also to Sidney Arenson, Peter Gumpert, Milton Rosenbaum, John Schopler, and Roger Wells for many helpful suggestions and critical readings of drafts of this report.

[2] Now at the University of Iowa.

are defined as tacit or explicit rules formed to influence the behavior of group members toward one another. Following Thibaut and Kelley (1959) and Thibaut and Faucheux (1965), norms are seen as impersonal substitutes for more direct means of interpersonal influence. Norms may be formed to maintain the integrity of the group during periods of interpersonal stress. Although the following analysis is confined to two-person interdependent groups, there is no reason that it could not be extended to larger groups.

In the interdependent dyad, each member has some control over the outcomes of the other (Thibaut & Kelley, 1959). Since the members of the dyad have power over one another, each can attempt to increase his own outcomes from the relationship by the use of his power to influence the other. Indeed, where the means of direct interpersonal influence are available, they are highly likely to be used. However, the injudicious use of punitive power and counterpower in an interdependent relationship can lead to disruption and consequent reduction of the outcomes of both members (e.g., Deutsch & Krauss, 1960, 1962). Thus it is useful to develop some adaptive mechanism to prevent such disruptive occurrences. Norms governing power use provide an appropriate device. Since norms are used to control behavior in a relatively depersonalized manner, appeals to them are not seen as assertions of personal superiority in the relationship, as appeals to punitive power (threats) often are.

If one person exercises power in an interdependent dyad, he implies by doing so that he occupies a superior position in the relationship and, as a by-product, probably angers the other as well as lowering his outcomes (Kelley, 1965). The other may respond by exercising counterpower in order to restore his sense of equal status in the relationship and to change the first person's behavior. Thus both members of the dyad are reduced to less than optimal outcomes by the other's action. The argument presented in this paper is that under certain circumstances the members of the dyad become sensitized to such an eventuality before it takes place and take preventive steps if such steps are available. More generally, if *both* members of the dyad perceive that the other member is tempted and able to exercise severe punitive power in the relationship, they will respond to this situation by attempting to agree, tacitly or explicitly, to refrain from exercising their power. Each has power which he can agree to abandon in exchange for security from the other's power. If this is accomplished by the formation of a publicly agreed upon norm, or *contract*, the control exercised through the norm appears to be extrapersonal and relatively nonthreatening to the harmony of the group.

Although it is not theoretically required, the members of the experimental dyads studied here were given different sources of power over each other. One member was given the responsibility of dividing the joint rewards of the relationship between himself and his bargaining partner. Although the dyad bargained over the division of rewards, the reward divider made the final decision about how they would be allocated. His partner's counterpower was based on his ability to leave the relationship for a somewhat desirable alternative; if he chose this alternative, the reward divider was left with his own, less desirable alternative relationship. Thus, the reward divider could enhance his own outcomes by exploiting his partner, provided he stayed in the relationship. The other person could be disloyal to the relationship and thus reduce the reward divider's outcomes to an unsatisfactory level. Each, then, had the potential to threaten or punish the other. Each member is probably unwilling to restrain his use of power unless the other has counterpower and seems likely to use it. It seems that contractual norms are most likely to be formed when each member perceives the other as likely to exercise his power. The kind of contracts the members are most likely to make should contain norms of loyalty and fairness.

In an experiment which tested the above formulation, the behavior of the partner was varied in terms of the fairness of the reward divider (A) and the loyalty of the other member (B). In an *unfair-A* condition, A was presented to B as tending to better his outcomes at B's expense, while in a *fair-A* condition, A was presented to B as not tending to better his outcomes at B's expense. In a *disloyal-B* condition, B was presented to A as tending to withdraw from the relationship, while in a *loyal-B* condition, B was presented to A as not tending to withdraw.

Smith (1965) has suggested that the overuse of punitive power is often seen by the user as leading to the loss of the good will of the person upon whom the power is used. Therefore, to avoid unintentional power overuse it is desirable for the user to have precise control of the degree of power he has over the other. This suggests that if the reward divider in the present experimental dyad had highly imprecise control over the use of his power his concern about the partner's possible withdrawal from the relationship would become greater. Moreover, under imprecise control conditions the partner should also be apprehensive about the overuse of punitive power against himself. Thus, the partner's fear of the extent of the reward divider's potential exploitativeness would also be increased under the conditions of imprecise power by the reward divider. It is suggested that the development of contractual norms would be more likely if the reward divider had imprecise control of his power than if his control was precise.

The precision of control available to A in the use of his power was manipulated in the present experiment by allowing him, in a *precise control* condition, to divide the joint rewards any way he wished, within a certain range of outcomes. In the *imprecise control* conditions, A was allowed only four divisions—the two at the extreme ends of the outcome range as in precise control, and two intermediate choices.

In summary, the above considerations were tested in a $2 \times 2 \times 2$ factorial design. The following predictions were made: (*a*) Contractual norm development is facilitated when each member of the dyad perceives that the other is tempted and able to use punitive power in the relationship; (*b*) imprecise control of punitive power leads to increased contract behavior compared with precise control of punitive power.

METHOD

The subjects, 192 male undergraduates, were recruited from an introductory psychology course.

Apparatus consisted of four adjoining booths, separated by 7-foot high partitions, arranged in semicircular fashion around a common front. A slot located in front of each booth permitted the sending and receiving of messages between subject and experimenter. From their booths, the subjects could not see each other or the experimenter.

The experimental bargaining game, which involved two players with different roles, was based on the game used by Thibaut and Faucheux (1965). The task of each player was to compete for points against a player who had his role in another dyad. In several plays of the game, each player independently decided to receive points from one of two point sources: the bargaining-choice point source (hereafter BC) or the alternative point source (AP). The BC represented the joint rewards of the relationship, while the AP represented rewards outside the relationship. The game roles gave the members of a dyad different kinds of power over each other. One member, A, was given the power to divide the points in the BC between himself and his partner. The other member, B, had counterpower via the AP. When B chose AP, each player received only the relatively low points in the alternative.

The details of the game are depicted in Table 1. The number of points in the BC and AP are shown in the note to the table, and the matrix indicates how these points were allocated depending on A's and B's game choices. Whenever A or B or both chose AP on game play, then only 66 points were available, 33 each for A and B. When A and B both chose BC on a play, the only points available for distribution on the play were the 98 in the BC source. As shown in Table 1, A could make any point split between 10 and 88 inclusive (A = 88, 87, ... 11, 10 / B = 10, 11, ... 87, 88). Actual game choices were made in private.

Procedure

An experimental session involved four subjects. The experimenter read them a standard set of instructions, which included the supposed aims of

TABLE 1

A'S GAME CHOICES AND POINT SPLITS[a]

B's game choices	AP	BC	. . .	BC	. . .	BC	. . .	BC
AP	33[b] 33	33 33	. . .	33 33	. . .	33 33	. . .	33 33
BC	33 33	88 10	. . .	54 44	. . .	44 54	. . .	10 88

Note.—Real-game chart used in the experiment in all treatments:

Bargaining Choice Point Source or BC	88/10 Total: 98
Alternative Point Source or AP	33/33 Total: 66

[a] In precise outcome control all intermediate cells were available. In imprecise outcome control all empty cells were unavailable.

[b] Numbers above the diagonal give A's points. Numbers below the diagonal give B's points.

the experiment, the general nature of the game and the competition, and how the game was played.

Two demonstration diagrams, shown one at a time, were used to describe the point sources and the role differences between A and B. The possible point outcomes to the members of a hypothetical A-B dyad were described. The demonstration games had dominant solutions, which ensured A and B maximum profits.

A booth was randomly assigned to each subject. Alphabetic letters in the booths indicated a player's role, partner, and opponent; each player thus knew the latter persons only by letter. An actual game diagram was in the booths, which subjects now saw for the first time (see note to Table 1).

The experimenter then introduced two incentives: an additional hour of credit for participation in experiments to that A and to that B of the two with more points at the end of the session, and a chance for the better A and B to win a further $20 lottery prize. There were separate lotteries for A players and B players.

The subjects were passed an information sheet which began the experimental manipulations. These sheets gave information about the game behavior of previous subjects:

1. In the unfair-A condition, Bs were told that most As favored a point split around 65 for A and 33 for B, etc.
2. In the fair-A condition, Bs were told that most As favored a point split around 49 for A and 49 for B, etc.
3. In the disloyal-B condition, As were told that most Bs tended to play AP about 5 out of 10 times, etc.
4. In the loyal-B condition, As were told that most Bs rarely played AP, etc.

Each sheet also gave some reasons why previous subjects consistently acted as they did. These sheets gave no information about others in their own role.

A practice was held. The subjects were passed a sheet on which they entered 10 practice game-play choices (and in the case of A, point splits) using the real-game chart. The practice sheets were returned to the experimenter, who said that an exchange of practice sheets between partners would be helpful in learning the game. The actual practice-game choices were not transmitted. The subjects received practice-sheet choices as follows: (*a*) In the unfair-A condition, A's 10 point splits gave B an average of 30 points per play; (*b*) in the fair-A condition, A's 10 point splits gave B an average of 48 points per play; (*c*) in the disloyal-B condition, B chose AP on 6 of the 10 plays; and (*d*) in the loyal-B condition, B chose AP on none of the 10 plays.

Precision of outcome control was manipulated at this stage by varying the number of steps in A's range of point divisions. The *precise* treatment, with its 79 possible point splits in the range 88/10, was described above. In the *imprecise* outcome control treatment, A was permitted to use only four splits in the 88/10 range (A = 88, 54, 44, 10 / B = 10, 44, 54, 88).

The subjects played two real-game plays with no communication by means of two game-choice sheets on which they filled out the same information as on the practice sheets. A check on the manipulation followed by asking the subjects to estimate their partner's choices on the two real-game plays. The As estimated how many times their B had played AP, and the Bs estimated the point splits made by their A.

After this, the experimenter introduced the information exchange between bargaining pairs. During the remaining game rounds, subjects in a pair had to decide among three courses of action about exchanges. They could (*a*) play as on the first two real-game plays, or (*b*) make suggestions about the game choice and/or point split, or (*c*) form contracts. A contract involved a (*a*) specified point split, (*b*) rule or rules (to be discussed below), (*c*) 7-point cost to each player, and (*d*) fine for violation of contract. The suggestions and contracts held for one play and had to be agreed upon by both players. Game choices for a play, made after the information exchange, were not communicated, and nothing was binding, not even contracts. However, when a player broke a contract rule, the amount of the fine (from 10 to 60 points) was deducted from the player's points.

The subjects received a written summary of the information exchange. It also introduced the rule(s) which had to be incorporated in a contract: forbidding the play of AP in private if it was agreed upon during the bargaining to play BC, and/or forbidding A to take more points in the private play than was agreed upon during the bargaining.

The information exchange was restricted to the sending and receiving of standard offers and messages prepared by the experimenter. The subjects wrote offers or messages by appropriately circling mimeographed sheets. Each player was restricted to three messages per game play.

After the fifth play, the subjects were informed of the deceptions used, and the purposes of the study were explained to them. They were requested not to discuss the experiment with other students. The four players at a session all won an extra hour of experimental credit. Their names went into the lotteries, which were held.

A total of 12 dyads were run in each of the eight conditions. The sessions (two dyads at a time) were run in predetermined combinations of conditions. Precise and imprecise outcome control alternated from one session to the next. Hence the same verbal instructions were used for all subjects in a given session. Second, dyads at the same session were under different conditions of fairness and loyalty.

The pairings of the two levels of fairness of A and loyalty of B will be referred to as follows: UA-DB (for unfair A-disloyal B), UA-LB (for unfair A-loyal B), FA-DB (for fair A-disloyal B), FA-LB (for fair A-loyal B).

RESULTS

Checks on the effectiveness of the manipulations of loyalty of B and fairness of A were

TABLE 2

INDEXES OF CONTRACTS IN THE FAIRNESS AND LOYALTY CONDITIONS SUMMED OVER OUTCOME CONTROL CONDITIONS

Fairness condition	Loyalty condition	
	Disloyal B	Loyal B
Unfair A		
No. dyads forming at least one maximally stringent contract	8	1
Mean fine for contract groups	52.5 (12)[a]	33.1 (8)
Fair A		
No. dyads forming at least one maximally stringent contract	1	2
Mean fine for contract groups	47.5 (6)	35.0 (7)

[a] The numbers in brackets indicate the number of dyads forming at least one contract.

respectively assessed by A's estimate of B's AP choices and by B's estimate of A's point splits. The data from As and Bs were analyzed in separate three-factor analyses of variance by experimental condition with 12 scores per cell. The within-cell distributions tended to be Poisson, so a square-root transformation was used. As anticipated, for Bs the mean point-split estimate in the unfair-A condition was greater than that in the fair-A condition. That is, there was a fairness of A main effect ($F = 31.3$, $df = 1/88$, $p < .001$), with no other statistically significant effects. For As there was a significant loyalty of B main effect on estimates of partner's AP choices. As desired, then, the mean AP estimate in the disloyal-B condition was greater than the mean AP estimate in the loyal-B condition ($F = 30.5$, $df = 1/88$, $p < .001$). In addition there was an unexpected triple interaction in this analysis which is discussed below.

The first hypothesis about contracts was that activity related to contracts should be significantly greater in UA-DB than in the other conditions. The results of two measures of contracts in UA-DB, UA-LB, FA-DB, and FA-LB summed over outcome control are summarized in Table 2.

TABLE 3

NUMBER OF MAXIMALLY STRINGENT CONTRACTS IN THE FAIRNESS, LOYALTY, AND OUTCOME CONTROL CONDITIONS

Fairness condition	Outcome control condition			
	Precise		Imprecise	
	Loyalty condition		Loyalty condition	
	Disloyal B	Loyal B	Disloyal B	Loyal B
Unfair A	5	0	3	1
Fair A	0	0	1	2

The most stringent contract that could be formed included both rules and a 60-point fine. Forming a maximally stringent contract indicates the most awareness of the sources of punitive power in the group. In UA-DB, 8 out of 24 dyads formed at least one maximally stringent contract during the three game plays when contracts were possible. The corresponding figures in UA-LB, UA-DB, and FA-LB, respectively, were 1, 1, and 2 dyads forming one or more such contracts. The Fisher exact test was used to make pairwise comparisons among the latter three conditions on number of groups forming at least one maximally stringent contract. All comparisons were statistically nonsignificant. However, the difference in contracts between the UA-DB and the other three conditions combined was significant ($p < .001$) by the chi-square test with the continuity correction. Thus, in accordance with prediction, the dyads in the UA-DB condition were more likely to form one or more maximally stringent contracts than the dyads of the other conditions.

TABLE 4

MEAN FINE FOR CONTRACT GROUPS IN THE FAIRNESS, LOYALTY, AND OUTCOME CONTROL CONDITIONS

Fairness condition	Outcome control condition			
	Precise		Imprecise	
	Loyalty condition		Loyalty condition	
	Disloyal B	Loyal B	Disloyal B	Loyal B
Unfair A	54.3 (7)[a]	33.3 (3)	50.0 (5)	33.0 (5)
Fair A	45.0 (2)	15.0 (2)	48.8 (4)	43.0 (5)

[a] See footnote in Table 1.

Another index of contracts was the amount of fine for breaking the terms of a contract. The analysis was restricted to groups forming one or more contracts. A t test was used to compare the mean fine per contract for the 12 contract groups in UA-DB with the 21 contract groups in the other three conditions. The respective means for the UA-DB and remaining conditions combined were 52.5 and 37.9 points ($t = 2.53$, $p < .02$). Thus the contracts formed in UA-DB were likely to include a higher fine for violation than those in the other conditions.

The second hypothesis about contracts was that there would be greater activity related to contracts under imprecise control of outcomes than under precise control of outcomes. Precise and imprecise outcome control conditions were compared by summing over fairness and loyalty conditions. Neither of the differences on the two dependent measures of contracts was statistically significant.

To understand what effects, if any, imprecise control of outcomes had on contract behavior, further data analyses were performed. A summary of the measures of contracts in the eight experimental conditions is presented in Tables 3 and 4.

Under precise (imprecise) control, UA-LB, FA-DB, and FA-LB were not expected to differ among themselves on the two dependent measures of contracts. Pairwise comparisons among UA-LB, FA-DB, and FA-LB were done separately for precise and imprecise control of outcomes. The Fisher exact-probability test was used to test differences on number of contracts. The regular t test, or the Welch procedure when variances were heterogeneous, was used to test differences on mean fine for groups who formed at least one contract. For precise (imprecise) control the six differences tested were statistically nonsignificant. The UA-DB precise (imprecise) was next compared with the remaining precise (imprecise) conditions combined. For precise control, UA-DB differed from the other three combined conditions on both number of contracts ($p = .0021$) and on mean fine ($p < .05$). The above findings are in line with prediction. However, for imprecise control, UA-DB was not statistically different from the remaining conditions on either index of contracts. Thus imprecise control, compared with precise control, leads to heightened contractual norm behavior in all but the UA-DB condition.

DISCUSSION

Two conditions were predicted to facilitate the formation of contracts in a dyad. One condition is the perception by both members that the other is likely to use punitive power in the relationship. Another condition is possession of crude outcome control by the reward divider of the relationship. Hypotheses relating to these conditions have received confirmation from the results of this study.

The effects on contracts of imprecise outcome control were not fully as predicted. Possession of imprecise control suggested overuse of punitive power, which was guarded against by the formation of contracts. This was found in three of the four cells under imprecise control when they were compared with the appropriate precise control cells (Murdoch, 1965). However, the UA-DB crude cell was atypical in not tending to exhibit higher contract indexes than the UA-DB precise cell.

In the section above on manipulation checks there was mentioned a Fairness × Loyalty × Outcome Control interaction. It occurred in the analysis of A's estimates of his partner's AP choices ($F = 8.32$, $df = 1/88$, $p < .01$). Duncan's new multiple-range test was applied to the mean AP choice estimates. A description of the interaction consistent with the test was that As in the UA-DB imprecise control cell did not tend to perceive their Bs as withdrawing as much as the As in the other three disloyal-B cells. Such a difference for the UA-DB subgroup could have led to fewer contracts in the imprecise control of outcomes UA-DB cell. Why the As may have perceived Bs differentially in this cell is another matter. The only suggestion is the negative one of sampling error.

When a member (P) of an independent dyad believes that the other is going to use his power in the relationship against P, conflict usually results (Kelley, 1965). The implication of the present research is that such conflicts may be resolved only when both members are tempted and able to use severe punitive power. The conflict is resolved through the development of an adaptive mechanism (contracts with appropriate rules).

REFERENCES

DEUTSCH, M., & KRAUSS, R. M. The effect of threat upon interpersonal bargaining. *Journal of Abnormal and Social Psychology*, 1960, **61**, 181–189.

DEUTSCH, M., & KRAUSS, R. M. Studies of interpersonal bargaining. *Journal of Conflict Resolution*, 1962, **6**, 52–76.

KELLEY, H. H. Experimental studies of threats in interpersonal negotiations. *Journal of Conflict Resolution*, 1965, **9**, 79–105.

MURDOCH, P. The development of contractual norms in the interdependent dyad with power differentiation. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill, 1965.

SMITH, W. P. Precision of control and the use of power in triads. Paper read at Eastern Psychological Association, Atlantic City, April 1965.

THIBAUT, J., & FAUCHEUX, C. The development of contractual norms in a bargaining situation under two types of stress. *Journal of Experimental Social Psychology*, 1965, **1**, 89–102.

THIBAUT, J. W., & KELLEY, H. H. *The social psychology of groups*. New York: Wiley, 1959.

(Received May 26, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 212–215

EFFECTS OF LOCUS OF CONTROL AND EXPECTATION OF FUTURE CONTROL UPON PRESENT PERFORMANCE[1]

DAVID WATSON AND EVELYN BAUMAL

University of Toronto

Female Ss, split on their scores on the Locus of Control scale, learned a list of paired-associate nonsense syllables either under instructions that those associates learned would later serve as avoidance responses, subject to their learning skill, or would sometimes serve as avoidance responses, to be determined by chance. Internal locus of control Ss made more errors when they anticipated not having control over the later situation, while external locus of control Ss made more errors when they anticipated possessing control over the later situation. The results are interpreted as indicating differential anxiety arousal in different situations, and a 2-variable theory of anxiety is proposed.

As measured on the Locus of Control (LC) scale (Rotter, 1966; Rotter, Seeman, & Liverant, 1962), an individual is said to have an internal locus of control (IC) if he usually appraises positive and negative reinforcements to be contingent upon his own actions and hence under his own control, while a person is said to have an external locus of control (EC) if he usually appraises reinforcements to be beyond personal control and not contingent upon his own actions. Cromwell, Rosenthal, Shakow, and Zahn (1961) found that schizophrenic patients who were significantly more EC than normals preferred and performed better on a task in externally controlled situations. Within a normal group, the preference and performance of subjects were related to their locus of control scores under autonomous (internal control) conditions. Rotter and Mulry (1965) reported a study in which IC subjects took more time to make a decision in a skill condition than in a chance condition, while EC subjects showed a reverse, albeit nonsignificant, effect.

These studies suggest that individuals perform most efficiently in situations in which the actual environmental locus of control and the person's preference for, or appraisal of, the locus of control are congruent. A congruent situation is one in which the person's expectation concerning the agency of control is confirmed. The expectation of IC persons is confirmed when they possess control, the expectation of EC persons is confirmed when control is external, or determined by chance. This study was designed to test the performance of individuals in congruent and incongruent situations.

The general hypothesis to be tested was that individuals in incongruent situations become anxious, and that this emotion interferes with their performance on complex tasks. The hypothesized anxiety would manifest itself by lowering the subjects' confidence, or by causing errors which would interfere with optimal performance.

Why should incongruity produce anxiety? One's appraised locus of control, the variable which the LC scale purportedly measures, may also be one's preferred locus of control. IC subjects may prefer internal control, and EC subjects may prefer external control. If one group of persons prefers internal control, while another prefers external control, this is presumably a reflection of their preference for situations which maximize their outcomes, that is, for situations in which they feel most comfortable. One group may have learned that optimal outcomes occur when control over reinforcement is internal, the other that optimal outcomes occur when control is external. If the maximal outcome in an avoidance situation is one which minimizes the anxiety attendant upon the avoidence, then subjects will be least anxious in that situation which represents their preference for either internal or external control. Thus, in a situation in which the outcome may be painful to the subject, and when the actual and preferred-appraised locus of control are incongruent, the subject will become anxious.

Sometimes the responses that the person is learning will benefit him in a future situation, but other times their future benefit is doubtful, or indeterminate. The question to be answered here is how do EC and IC persons perform when a response they are presently learning may or may not serve to obtain some desired outcome later? If a response may or may not serve in the future as an avoidance response, will there be differences in its acquisition between IC and EC individuals? In line with the reasoning above, IC subjects would be in the preferred situation when their own learning skill determined their outcome, and hence they would not become too anxious, while EC subjects would be in the preferred situation when chance determined their outcome and hence would not become too anxious. However, IC subjects would become anxious and thus not perform well when chance determined their outcome, and EC subjects would fail to perform well when their own skill determined their outcome.

---

[1] This study was supported by Grant 67 from the Ontario Mental Health Foundation.

This experiment was designed to observe the performance of IC and EC subjects in learning responses that could produce a desired outcome—avoiding electric shock—when either learning skill or chance determined the outcome. Given experimental situations in which either skill or chance determine the avoidance of a noxious shock, the acquisition performance should reveal interaction effects. IC subjects will be less anxious and thus perform better under skill conditions, where the actual and preferred-appraised locus of control are congruent, while EC subjects will be less anxious and thus perform better under chance conditions, which is for them a congruent situation.

METHOD

Subjects and Task

The LC scale was administered to 565 females in introductory psychology at the University of Toronto. From the top 10%, EC subjects with scores of from 15 to 21, and from the bottom 10%, IC subjects with scores of from 1 to 4, 60 subjects were randomly selected, 30 from each end of the distribution. Within each group of 30 subjects, assignment to either the skill or chance conditions was random. Thus the design of the study was a 2 × 2, IC or EC subjects and skill or chance outcome; total $N = 60$, 15 in each cell.

The subject's task was to learn a list of eight paired-associate nonsense syllables. The pairs constructed were of 53% association value. They were DUP-ZET, BOZ-XUN, NIC-KAW, GEV-SOH, FAX-HUJ, RIQ-CEG, WOB-JIR, and VAY-NOK. The syllables were presented in five different orders. They appeared on a memory drum every 3 seconds with a 3-second exposure. The stimulus syllable appeared first, followed by the pair. The subjects were asked to correctly anticipate the response to the stimulus syllable before the pair appeared.

Procedure

Subjects were tested individually by a female experimenter. At the beinning of the experimental period, subjects were informed that they were participating in an experiment in avoidance learning. They were first to learn a list of paired nonsense syllables. They would then be given a second list to learn. The second list would consist both of pairs which they had already learned and of new pairs. Up to this point instructions were the same for all subjects.

Those in the skill condition were then informed that, while learning the second list, stimuli which they had already learned in the first list would sometimes appear, and if they were unable to provide the correct response, they would receive a moderately painful electric shock. Thus they would be punished for incorrect responses or failures to respond. Whenever they could anticipate the correct response, however, they would avoid the shock. Thus avoidance of shock would depend upon their ability to make a response which they had already learned. This procedure was to continue until they had learned the second list.

Those subjects in the chance condition were given similar instructions; they would receive the moderately painful shock when they could not correctly anticipate responses learned on the first list. However, their avoidance of the shock would be on a chance basis: they might or might not be shocked even when they did correctly produce the response. If the subject could not produce the correct response she would definitely be shocked, and if she did produce the correct response she might or might not be shocked, this to be determined on a chance basis. Actually, no subjects were shocked, as the second list was never presented.

Following the instructions, subjects began the task of learning the paired-associate lists. The subject read the stimulus syllable when it appeared, attempted to anticipate the response syllable, and then read the pair when both appeared. The list was continuously presented until the subject reached a criterion of one successful repetition of the complete list.

The number of errors, including incorrect responses and omissions, and the number of trials to criterion were recorded by the experimenter. When the subject reached criterion, the experimenter handed her a questionnaire to answer. The items were: (*a*) How anxious do you feel about learning the second list? (*b*) How confident do you feel that you will be able to anticipate the correct nonsense syllables for those pairs that are from List 1? The subject responded to these questions on a 10-point scale labeled "not at all" and "very" at the extremes. Subjects were then asked by the experimenter,

> Before you start on the second list, you may wish to practice going through the first list for some more trials. You can have from zero to ten trials to go through the first list. How many would you like?

The subjects' answers were recorded. It was felt that those subjects who were apprehensive about their performance would request more practice trials. The study was completed at this point by telling the subject that there would be no second list, explaining the study to her—except that the LC measure was not mentioned—and asking her not to reveal its nature to other students.

RESULTS

The number of errors and the number of trials to criterion were employed as dependent variables. It was felt that the more anxious the subjects were, the more errors they would make, and consequently, the more trials to criterion would be required. Table 1 presents the means for the errors analysis. As hypothesized, there was a significant interaction effect: IC subjects made more errors in the chance condition than in the skill condition, yet EC subjects showed the reverse effect ($F_{1,56} = 4.37$, $p < .05$). There were no other significant effects. These results may imply less anxiety for IC subjects in the skill condition and for EC subjects in the chance condition.

The analysis for trials to criterion, which is obviously correlated with errors, showed a similar interaction effect which approached significance ($F_{1,56} = 3.16$, $p < .10$), the IC subjects requiring fewer trials to learn the list in the skill condition, while the EC subjects required fewer trials in the chance condition.

Table 2 presents the mean number of practice trials requested by the subjects. It was expected that subjects who felt unsure of being able to produce the correct avoidance response, either because they were in what was to them an incongruent situation or because they had not learned the responses well, would want to have more practice trials. A significant interaction effect was observed ($F_{1,56} = 9.19$, $p < .01$), but there were no other significant effects. The IC subjects requested fewer practice trials in the skill condition, while the EC subjects requested fewer trials in the chance condition. The interaction pattern is identical to that observed in the analysis of errors.

Phares (1962), who did not use an individual difference measure, reported parallel results from a similar experiment: subjects who expected to be able to escape from shock on a skill basis showed heightened perceptual vigilance.

TABLE 1

MEAN NUMBER OF ERRORS TO CRITERION

Control	Skill	Chance
IC	46.5	65.9
EC	60.5	49.2

TABLE 2

MEAN NUMBER OF PRACTICE TRIALS REQUESTED

Control	Skill	Chance
IC	3.7	5.3
EC	6.5	4.0

Analysis of the subjects' responses to the post-task question which asked how confident they were about their performance on the second list revealed a significant difference between the IC and EC groups ($F_{1,56} = 4.90$, $p < .05$). The means were IC, 6.0 (out of 10), EC, 4.8, indicating that the IC subjects were more confident. This suggests that there is some relationship between the LC scale and self-esteem or anxiety variables. Indeed, one of the authors (Watson, 1966) reported a correlation of .36 ($N = 648$, $p < .01$) between the LC scale and manifest anxiety.

The difference between the skill and chance conditions almost reached significance ($F_{1,56} = 3.82$, $p < .10$). The means were: skill, 5.6; chance, 4.9. The subjects might have felt more confident in the skill condition because they knew it was possible to avoid the shock completely in that condition.

The analysis of the subjects' responses to the question, "How anxious do you feel about learning the second list?," did not produce any significant effects. In retrospect, it can be seen that this was the wrong question to ask. The subjects should have been asked how anxious they felt while learning the first list. There is little reason to expect that they should be anxious or nervous about learning the second list; nothing is contingent upon their good performance on that list.

DISCUSSION

The IC subjects in the chance condition and the EC subjects in the skill condition made more errors and requested more practice trials. If anxiety is conceived as a nonfacilitative level of motivation, then these results confirm our hypotheses. Under certain conditions the subjects become anxious and hence make errors in learning the list. The fact that there were no differences between the groups in their response to the questionnaire item asking them how anxious they would be calls an anxiety interpretation of the results into question. It was noted, however, that the wrong question was asked, and that there was in fact little reason to expect differences in response to the question. Yet it is possible that the subjects made more errors not because they were anxious, but simply because they were not motivated. This interpretation is belied by the fact that the same subjects requested more practice trials; surely if they were simply unmotivated they would not have wanted more practice. We conclude, therefore, that the most parsimonious explanation of these results is that under certain conditions different groups of subjects become anxious.

It appears that IC subjects are more concerned about their performance when they are in situations in which they do not have control; that is, they make more errors and want more practice in such situations, while EC subjects show more concern in situations in which they do possess control. Further, IC subjects are more confident of ability than are EC subjects. The EC subjects are implicitly saying, in this experiment, that they do not have a very high regard for their own ability to perform successful avoidance responses, and that they prefer control to be external to them, while IC subjects seem to be saying the opposite. The LC scale seems to measure preference as well as appraisal, and the EC part of the distribution measures a variable correlated with anxiety and self-esteem.

Rotter and Mulry (1965) noted that IC and EC subjects "differ in the value placed on the same reward depending upon whether it is perceived as contingent upon chance or skill [p. 598]." In their study IC subjects took longer to make decisions under skill instructions, while in the present study they made fewer errors under such instructions. EC subjects showed reversed effects. It has been suggested that the present results can be explained by appeal to a mediating anxiety variable.

The parallel between our results and Rotter and Mulry's suggests a two-variable explanation of the occurrence of anxiety: the outcome must be valued by the subject, and the means of attaining the outcome must be incongruent with the subject's preferred and expected means. Anxiety occurs when the subject's expectations and preferences concerning means to ends are not met and when the ends are highly valued ones. Mandler and Watson (1966) offer a similar explanation for the occurrence of anxiety. They state that anxiety occurs when an organized response sequence is interrupted, and there is no alternate response sequence available.

REFERENCES

CROMWELL, R., ROSENTHAL, D., SHAKOW, D., & ZAHN, T. Reaction time, locus of control, choice behavior and descriptions of parental behavior in schizophrenic and normal subjects. *Journal of Personality,* 1961, **29**, 363–380.

MANDLER, G., & WATSON, D. Anxiety and the interruption of behavior. In C. Spielberger (Ed.), *Anxiety and behavior.* New York: Wiley, 1966. Pp. 263–288.

PHARES, E. J. Perceptual threshold decrements as a function of skill and chance expectancies. *Journal of Psychology,* 1962, **53**, 399–407.

ROTTER, J. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs,* 1966, 80(1, Whole No. 609).

ROTTER, J., & MULRY, R. Internal versus external control of reinforcement and decision time. *Journal of Personality and Social Psychology,* 1965, **2**, 598–604.

ROTTER, J., SEEMAN, M., & LIVERANT, S. Internal versus external control of reinforcement: A major variable in behavior theory. In N. Washburne (Ed.), *Decisions, values, and groups.* Vol. 2. New York: Pergamon Press, 1962. Pp. 473–516.

WATSON, D. Relationship between locus of control and anxiety. *Journal of Personality and Social Psychology,* 1967, **6**. 91–93.

(Received June 3, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 216–221

A STUDY IN TIME ORIENTATION:

THE RELATION BETWEEN MEMORY FOR PAST EXPERIENCE AND ORIENTATION TO THE FUTURE[1]

JUDITH MARCH GOLDRICH
New York University

An assumption was made that time orientation to the past and time orientation to the future are predictably interrelated. If the individual remains involved with aspects of past experience, it seems likely that he will have relatively less energy available for the pursuit of future goals. An avowed recall scale was used to measure past orientation with a sample of 80 doctoral students. ½ of the sample had been judged by faculty to be unusually efficient and ½ unusually inefficient in promoting their own progress toward dissertation completion. As predicted, efficiency was found to be related to a median degree of avowed recall and inefficiency to the extremes of recall. In accordance with expectation, the efficient group and the median recall group introduced future referents significantly more often on the TAT than did their peers. Efficient subjects demonstrated greater awareness of the continuity of events in time, more involvement in their productions, and less tendency to describe unpleasant events. This group also made higher scores on a scale of conscious optimism about the future. Significant differences found between all groups in verbal productivity on the TAT lent some support to the original energy hypothesis.

The concept of time orientation as a dimension which contributes significantly to the understanding of the individual (both alone and as a member of a group) has long been proposed by scientists with diverse interests in the field of human behavior (Balken, 1943; Lewin, 1942, 1946; Oberndorf, 1941; Sturt, 1925). Anthropologists and sociologists have shown that cultures, at various phases in their existence, as well as various social classes within one culture, tend to entertain temporal orientations unique to themselves, which serve to shape, organize, and control communal activities (Arieti, 1947; Ketchum, 1951; Smith, 1952; Werner, 1948). Russell takes the view that civilizations mature in a manner parallel to individuals, by gradually abandoning activities which are oriented toward short-term impulse gratification and taking up instead functions which demand foresight, delay, and synchronization of group activities, that is, a future-oriented time sense (Arieti, 1947).

In a study with significant implications for education, Leshan (1952) demonstrated differences in temporal attitudes of lower- and middle-class children. Through use of a story-telling technique, he found that lower-class children characteristically view the world in terms of shorter time spans, with sharply limited periods of gratification delay. They plan ahead far less than their middle-class peers and respond less to threats of punishment and promises of reward which are localized in the future.

The individual's transition from behavior dominated by the urgency of impulses (pleasure principle) to behavior of a goal-directed and therefore future-oriented nature (reality principle) is described by Freud (1949) in his discussion of the processes of psychic maturation. Adler, and later the existential psychoanalysts, have gone so far as to state that they consider the degree and quality of the individual's future orientation to be the measure par excellence of his mental health (Adler, 1925; Allport, 1950; May, 1958; Way, 1950).

Some support for this view may be found in repeated demonstrations that people who are emotionally disturbed and/or socially delinquent tend to maintain distinctive and distorted time perspectives whose common denominator is a defective, limited future orientation (Federn, 1952; Fenichel, 1951; Israeli, 1936). Clinical observers have noted for example that depressives behave as though their temporal field had shrunk to exclude the future (Minkowski, 1958). Escalona (1940) found experiences of success and failure to have virtually no effect on level of aspiration in depressives. Similar deficits in future perspective have been demonstrated in delinquent children, acting-out adolescents, psychopaths, and schizophrenics (Balken, 1943; Barndt & Johnson, 1955; Davids, Kidder, & Reich, 1962; Greenacre, 1945; Redl & Wineman, 1951). In an investigation of factors prognostic of recovery in hospitalized schizophrenics, Levine [2] concluded that there was greater predictive value in measures of the patients' attitudes toward the future than in indicators of symptom severity.

---

[1] This research was supported by a predoctoral fellowship from the National Institute of Mental Health and was the basis of a doctoral dissertation submitted to New York University. The author wishes to express her gratitude to Walter Wilke for his assistance and advice.

In the realm of personality theory, variations in the time perspective and especially in the strength of future orientation have been consistently associated with variations in the achievement motive (McClelland, Atkinson, Clark, & Lowell, 1953). Investigators have been able to demonstrate that individuals who make high need-achievement scores tend very much to be "time-driven" people, to perceive events as being in rapid motion, and to strive actively for control and management of their time (Green & Knapp, 1959; Knapp, 1962; Knapp & Garbutt, 1958). Significant correlations have been demonstrated between extension of future perspective in projective material and academic achievement in grade and high school when the factor of intelligence is controlled (Davids & Sidman, 1962; Teahan, 1958). Using Harvard freshmen as subjects, Ricks and Epley (1960) demonstrated the same relationship. It was part of the intention of this study to determine whether similar temporal measures can discriminate effectively, on a higher level of academic achievement, between doctoral students who show efficient or inefficient progress toward completion of dissertation work.

Past investigations of time orientation have generally focused on future perspective and have not explored the relationships which may exist between the individual's orientations to various time dimensions including past and present. In this study, the assumption was made that time orientation is a *general function* and that associations may exist between its various components.

Assuming that an individual has at his disposal a finite amount of energy, it seemed logical to hypothesize that the lesser his tendency to remain involved with past experience, the greater were his chances of efficient successful progress toward attainment of future goals. Since recollection is a major avenue of contact between the individual and his past, it was reasoned that a good index of continuing involvement with the past would be relative amount of *avowed recall for small details of past experience.*

High degrees of retention (relative to scores of a homogeneous group of subjects) seem on the surface to indicate greater than average interest in the past. But low avowed recall may reveal equally great expenditures of energy in attempts to cope with the past, with the difference that these attempts are operating in the direction of warding off contact with memory. The rationale was therefore that either extreme in the recall function generally represents continuing involvement with the past and is indicative of a temporal orientation which is expected to be inimical to effective progress toward future objectives. It should be stated at this point that such a measure runs the risk of confounding the time-orientation dimension with other issues, for of the various motivations for recollection many exist, which do not connote overinvolvement with the past; to name one, recollections form a large part of the individual's sense of identity. However, it was felt that, by and large, people who avow the extremes of high or low recall for material from the past are displaying a temporal orientation which is significantly different from that of the median range recaller.

The *general hypothesis* deriving from these theoretical assumptions was the following: Individuals who avow either extreme of high or low recall for small details of personal history can be shown to be less capable of coping constructively with the future than are those whose orientation to the past is relatively free of defensive maneuvers and whose memory, consequently, is likely to be midway between that of the high and low recallers.

METHOD

Subject Selection

At two New York City universities, several graduate departments (all in the humanities) were selected which offered students approximately similar conditions for the pursuance of their dissertation work. The chairmen or departmental advisors served as judges for purposes of subject selection.

Lists of potential subjects included all students who were at the stage in their academic careers at which the average student in that department was already embarked on dissertation work. The age range was from 25 to 40. Judges classified the students on the basis of their subjective experience, stemming from years of acquaintance with work progress in the department, as either *efficient, average,* or *inefficient* in dissertation activities. Judges were requested to strike from the lists those students in whose lives there were known situational factors which might limit or control the rate of their work.

---

[2] D. Levine, "Rorschach Genetic Level and Psychotic Symptomatology," unpublished manuscript, 1960.

TABLE 1

PERFORMANCE OF THE PROGRESS GROUPS ON THE MEMORY SCALE

	Efficient *Ss*	Inefficient *Ss*
MS	34.40	36.62
Variances	22.94	39.43*

* $F = 1.7190$, $p < .05$.

The *average* group was discarded in order to sharpen the progress measure. Subjects were told that the study involved validation of a new psychological measure with an intellectually superior population. The sample consisted of 40 *efficient* and 40 *inefficient* subjects. In the first group, there were 26 men and 14 women; in the second group, 31 men and 9 women. Sex effects were extracted in statistical computations.

Procedure

Subjects were individually tested on the following measures:

1. *TAT Cards 1, 2, 20, and 16*. Standard instructions were used with all temporal references deleted, in order to avoid the possibility of introducing the individually varying and uncontrolled influence of authority orientation into test responses. Appropriate inquiry ascertained the length of temporal dimensions in subjects' responses.

2. *Memory Scale*. This was an original scale of 50 items which elicited avowed recall for small details of remote and recent past experience. Certain items lent themselves to estimations; therefore, three response alternatives were provided: answer, estimation or guess, and do not know. Illustrative items were as follows: "What was the number of the last telephone call you made and to whom was it made?" "How tall were you when you were 13 years old?"

3. *Future Scale*. An original scale of 20 Likert-type items was designed to measure an individual's conscious feelings, be they optimistic or pessimistic, about his future professional life, interpersonal relations, and self-evaluations, as well as his affective reactions to anticipations of the future. As an example, "Do you expect, in the course of your career, to have an opportunity to do the kind of work you enjoy?" The TAT was presented first to avoid contamination of projective responses by previous references to past and future.

Hypotheses

The progress measure (inherent in the method of subject selection) was chosen to reflect subjects' ability to promote their progress from the present status of student to the desired future status of qualified professional. Since completion of the doctorate is an acknowledged goal of all doctoral students, and since they are assumed to function on intellectual levels sufficiently high to support attainment of the goal, the assumption was made that variations in work efficiency reflected individual differences in readiness and/or capacity to move into the future in this particular area. This assumption led to the following hypotheses concerning the role of time orientation:

1. Those subjects judged to be *efficient* in dissertation work would tend to display a median degree of avowed recall, relative to the total sample, for small details of past experience, while subjects judged *inefficient* in dissertation progress would tend to show either high or low avowed recall.

2. Subjects displaying a median degree or avowed recall would exceed other subjects in the following aspects of their TAT productions: (*a*) number of stories in which the future is spontaneously mentioned, (*b*) prospective (future) time spans, (*c*) total time spans, (*d*) percentage of verbal output devoted to descriptions of future events, (*e*) continuity of depicted events in time, (*f*) pleasantness of stories, (*g*) degree of involvement of the author in his stories. Subjects displaying high or low recall would exceed other subjects in the description of unpleasant or frustrating events.

3. *Efficient* subjects would exceed the *inefficient* subjects on those dimensions of the TAT covered above under Hypothesis 2, *a* through g. Similarly, the *inefficient* subjects were expected to exceed in the description of unpleasant events in their stories.

Scoring

Blind scoring of TAT protocols was performed independently by three clinical psychologists who evaluated the four sets of responses separately. Scores were assigned on appropriate scales presented to the judges. In the case of dimensions *e* through *g*, as well as "description of unpleasant events," a clinical judgment was required and was rendered by reference to a simple 3-point scale.

Scores for the Memory Scale and the Future Scale required merely tallying of assigned values for the responses. In the case of the Memory Scale, inasmuch as estimates are in some cases as appropriate as answers for measuring avowed recall, and previous research has demonstrated that men tend to guess significantly more often on this scale than women, the total score consisted of total number of answers plus total number of guesses or estimations.[3]

RESULTS

The reliability of the Memory Scale was .76 for this sample (Kuder-Richardson 20). Hypothesis 1, which predicted a significant divergence in the variance of memory scores between *efficient* and *inefficient* subjects, was borne out at the .05 level on a one-tailed test (*F* test for the contrast of two variances). An inspection of Table 1 shows that the mean scores of the two groups do not differ significantly.

---

[3] J. March, "Time Orientation and Memory for the Past," unpublished manuscript, 1960.

To test the predictions concerning TAT responses of median versus extreme memory groups, the obtained distribution of memory scores was divided into equal thirds of the scale between the highest and the lowest scores. Done in this fashion, the high memory group consisted of 18 subjects, the median group of 42 subjects, and the low group of 20 subjects.

Interjudge reliability on the TAT ranged from .78 to .99 with a median of .97 (variance form). (Since many of the scores were objective, subject having specified the length of the temporal dimensions in his stories, agreement was naturally high.) Hypothesis 2 predicted a curvilinear relationship between memory-group membership and TAT scores; thus statistical treatment was by an analysis of various testing for the significance of both linear and curvilinear components in the regression. Inspection of Table 2 indicates that two of the eight predictions subsumed under this hypothesis were borne out at recognized significance levels, while in two other instances results were decidedly in the predicted direction. All other findings were nonsignificant. There were no significant linear components.

TABLE 2

RELATIONSHIP BETWEEN MEMORY-GROUP CLASSIFICATION AND TAT SCORES

TAT dimension	F	
	Curvilinearity	Linearity
Number of spontaneous future mentions	4.6928*	.1955
Prospective spans	3.5126	.2269
Total time spans	.3859	1.3812
Percentage of words devoted to future events[a]	7.3131**	.5736
Continuity in time	.1853	.0156
Pleasant events	.4744	.3054
Involvement in stories	1.8195	1.3166
Unpleasant events	.8980	.1877

[a] Percentages are transformed into arc sins to effect constant variance.

* $p < .05$.

** $p < .01$.

To test the last prediction concerning the relation between TAT scores and efficiency in dissertation work, statistical treatment was done by an analysis of variance with corrections for disproportionate cell frequencies, by the method of unweighted means. As the results presented in Table 3 indicate, the hypothesized relation was shown to exist at significant levels in three cases (Hypothesis 3). All other findings, with the exception of description of "pleasant events," which did not discriminate between groups, were in the predicted direction. There were no significant sex or interaction effects in TAT scores.

TABLE 3

RELATIONSHIP BETWEEN PROGRESS CLASSIFICATION AND TAT SCORES

TAT dimension	Direction of difference	Significance level
No. spontaneous future mentions	Predicted	.04
Prospective spans	Predicted	.25
Total time spans	Predicted	.25
Percentage of words devoted to future events[a]	Predicted	.25
Continuity in time	Predicted	< .005
Pleasant events	None	*ns*
Involvement in stories	Predicted	.01
Unpleasant events	Predicted	.08

[a] Percentages are transformed into arc sins.

An unpredicted finding related to total verbal productivity on the TAT. Inspection of Table 4 shows that the *efficient* subjects were significantly more talkative than the *inefficient* subjects and that the median memory group similarly significantly exceeded the extreme groups in total words produced.

A measure of the "percentage of words used to describe present events" also showed group differences. Scores of the *inefficient* group tended to exceed those of the *efficient* group ($p < .10$). When subjects were classified according to memory groups, the analysis of variance showed a strong curvilinear component in the regression ($F = 5.5164$, $p < .03$); however, the curve associated with these data was opposite to that which was most characteristic of the findings—the median group made lower scores on this dimension than the extreme groups.

TABLE 4

DIFFERENCES IN TOTAL WORD PRODUCTION ON THE TAT

Group	Mean no. words produced	Statistical findings
Progress		
Efficient	253.92	Analysis of variance, F (1, 76) = 4.1825, $p < .05$
Inefficient	201.70	
Memory		
Low	170.70	Analysis of variance testing for linearity & curvilinearity; curvilinear component, F (1, 77) = 8.3637, p = .005
Median	260.21	
High	215.89	

The reliability of the future scale was found

TABLE 5

ANALYSIS OF VARIANCE FOR THE RELATION BETWEEN CONSCIOUS OPTIMISM, PROGRESS CLASSIFICATION, AND SEX

Source of variation	df	*SS*	*MS*	*F*
Efficiency (A)	1	410.80	328.3768	3.5164
Sex (B)	1	648.01	565.5868	6.0566**
A × B	1		4.5205	.0484
Individual	76	7097.14	93.3832	

** $p = .02$.

to be .86 (variance form). Results of the analysis of variance are presented in Table 5. Both sex and progress classification were significantly related in performance on this measure. The *efficient* students expressed a greater degree of conscious optimism about the future than did the *inefficient* subjects, and the women were more optimistic than the men. No significant associations existed between scores on the memory scale and those on the future scale.

DISCUSSION

A significant association has been demonstrated between extremes of avowed recall for details of the past and inefficiency in the attainment of an important future goal, the doctoral degree, or conversely, between efficiency in this behavior and a moderate degree of such recall. In addition, the data show that the median memory group distinguished itself from the high and low groups by a significantly greater tendency to introduce future referents into projective responses.

The lack of relation between scores on the future scale and the memory measure does not bear directly on our hypotheses, for the association posited here is between "orientation" to past and future, actual behavior vis-à-vis these time dimensions in the individual's life rather than stated attitudes toward them. It is possible that the significant relation which was found between progress classification and future scores indicates that optimism is a predictor of future orientation; however, it seems as likely that the relative proximity of the two groups of subjects to their academic goal determined the divergence in their optimism.

Hypothesis 3 attempted to relate two measures of future orientation to one another without implying a causality which would seem premature in a relatively young area of research. Differences in TAT scores between *efficient* and *inefficient* subjects emerged in the predicted direction on seven of eight dimensions, but attained significance levels of .08 or less in only four instances.

It is not surprising to find that students judged *efficient* in dissertation progress displayed significantly greater involvement in their projective responses than did the *inefficient* group. They also demonstrated greater awareness of the continuity of events in time. It is suggested that one important component of the ability to pursue goals successfully may be concrete awareness of the interrelated steps which bridge the gap between present and future situations, between formulation and attainment of a goal. If *inefficient* subjects produce stories in which events succeed one another abruptly or magically or without logical connecting links, might they not prove equally unrealistic in approach to a complex academic project?

Researchers and psychoanalytic writers have stated that willingness to deal with the future represents a healthy "openness" to experience (Adler, 1925; Barndt & Johnson, 1955; Federn, 1952; Oberndorf, 1941). Opposite to this openness, in terms applicable to the TAT, would seem to be a "constriction" of the time dimension, or emphasis upon the present to the relative or complete exclusion of past and future. In this context, the finding that the *inefficient* group and the extreme memory groups tend to speak more about the present in their productions than do their peers seems congruent with the theoretical assumptions of the study.

Similarly, review of the literature discloses that most investigators attribute discrepancies in amount of verbalization on the TAT, such as those found here, to variations in drive level (Lindzey & Silverman, 1959; MacBrayer, 1959; Milam, 1954; Solkoff, 1960; Welch, Schafer, & Dember, 1961). This would suggest that those students who fell into the *efficient* and median memory groups possessed higher inherent levels of available energy than the others. Such an interpretation would lend support to the assumption that individuals making extreme memory scores have more energy invested in the past and less available for the pursuit of future goals than do those in the median group. However, experimental evidence is somewhat contradictory on this score, and further study is required to clarify the interpretation of this finding (McClelland, et al., 1953).

REFERENCES

ADLER, A. *Individual psychology.* London: Routledge & Kegan Paul, 1925. Pp. 100–106.

ALLPORT, G. W. *The individual and his religion.* New York: Macmillan, 1950. Pp. 125–131.

Arieti, S. The processes of expectation and anticipation. *Journal of Nervous and Mental Disease,* 1947, **106**, 471–481.

Balken, E. R. A delineation of schizophrenic language and thought in a test of imagination. *Journal of Psychology,* 1943, **16**, 239–271.

Barndt, R. J., & Johnson, D. M. Time orientation in delinquents. *Journal of Abnormal and Social Psychology,* 1955, **51**, 343–345.

Davids, A., Kidder, C., & Reich, M. Time orientation in male and female juvenile delinquents. *Journal of Abnormal and Social Psychology,* 1962, **64**, 239–240.

Davids, A., & Sidman, J. Impulsivity, time orientation and delayed gratification in future scientists and in under-achieving high school students. *Exceptional Children,* 1962, **29**, 170–174.

Escalona, S. The effect of success and failure upon the level of aspiration and behavior in manic-depressive psychoses. *Iowa University Studies in Child Welfare,* 1940, **16**, 199–302.

Federn, P. *Ego psychology and the psychoses.* New York: Basic Books, 1952.

Fenichel, O. On the psychology of boredom. In D. Rapaport (Ed.), *Organization and pathology of thought.* New York: Columbia University Press, 1951. Pp. 349–361.

Freud, S. Formulations regarding the two principles in mental functioning. (Orig. publ. 1911) In E. Jones (Ed.), *Collected papers.* Vol. 4. London: Hogarth Press, 1949. Pp. 13–21.

Green, H. B., & Knapp, R. H. Time judgment, aesthetic preference, and need for achievement. *Journal of Abnormal and Social Psychology,* 1959, **58**, 140–142.

Greenacre, P. Conscience in the psychopath. *American Journal of Orthopsychiatry,* 1945, **15**, 495–509.

Israeli, N. *Abnormal personality and time.* New York: Science Press, 1936.

Ketchum, J. D. Time, values and social organization. *Canadian Journal of Psychology,* 1951, **5**, 97–109.

Knapp, R. Attitudes toward time and aesthetic choice. *Journal of Social Psychology,* 1962, **56**, 79–87.

Knapp, R., & Garbutt, J. T. Time imagery and the achievement motive. *Journal of Personality,* 1958, **26**, 426–434.

LeShan, L. Time orientation and social class. *Journal of Abnormal and Social Psychology,* 1952, **47**, 589–592.

Lewin, K. Time perspective and morale. In G. Watson (Ed.), *Civilian morale.* New York: Houghton Mifflin, 1942. Pp. 48–70.

Lewin, K. Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946. Pp. 918–970.

Lindzey, G., & Silverman, M. TAT: Techniques of group administration, sex differences and the role of verbal productivity. *Journal of Personality,* 1959, **27**, 311–323.

MacBrayer, C. T. Relationship between story length and situational validity of the TAT. *Journal of Projective Techniques,* 1959, **23**, 345–350.

May, R. Contributions of existential psychotherapy. In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence.* New York: Basic Books, 1958. Pp. 37–92.

McClelland, D., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Milam, J. R. Examiner influences on TAT stories. *Journal of Projective Techniques,* 1954, **18**, 221–226.

Minkowski, E. Findings in a case of schizophrenic depression. In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence.* New York: Basic Books, 1958. Pp. 127–139.

Oberndorf, C. P. Time—its relation to reality and purpose. *Psychoanalytic Review,* 1941, **28**, 139–155.

Redl, F., & Wineman, D. *Children who hate.* Glencove, Ill.: Free Press, 1951. Pp. 116–128.

Ricks, D., & Epley, D. Foresight and hindsight in the TAT. Paper read at Eastern Psychological Association, New York, April 1960.

Solkoff, N. Effects of a variation in instructions on responses to TAT cards. *Journal of Projective Techniques,* 1960, **24**, 67–70.

Smith, M. W. Different cultural concepts of past, present and future. *Psychiatry,* 1952, **15**, 395–400.

Sturt, M. *The psychology of time.* New York: Harcourt, Brace, 1925.

Teahan, J. B. Future time perspective, optimism, and academic achievement. *Journal of Abnormal and Social Psychology,* 1958, **57**, 379–380.

Way, L. *Alfred Adler's place in psychology.* London: Allen & Unwin, 1950. Pp. 73–82.

Welch, B., Schafer, R., & Dember, C. TAT stories of hypomanic and depressed patients. *Journal of Projective Techniques,* 1961, **25**, 221–232.

Werner, N. Primitive notions of time. In, *Comparative psychology of mental development.* New York: International Universities Press, 1948. Pp. 182–190.

(Received June 6, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 222–224

AUTHORITARIANISM AND SANCTIONING BEHAVIOR[1]

DAVID S. DUSTIN AND HENRY P. DAVIS

University of Texas

40 undergraduate men—20 in the upper and 20 in the lower 4/10 of a distribution of balanced authoritarianism (F) scores—administered sanctions to a mythical follower after each of 15 follower-performance trials. In their attempts to maximize follower performance, Ss could use monetary rewards or penalties, written notes indicating positive or negative evaluations, or no sanction at all. Low-F Ss used negative sanctions (monetary penalty or negatively evaluative note) significantly less often than did high-F Ss, as expected, and used no sanction at all significantly more often than did the highs.

Among the characteristics of the authoritarian personality described by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) is a preference for certain ways of sanctioning a subordinate. This hypothesized characteristic is described most clearly in the following passage:

> The authoritarian must maintain strict discipline if he is to evoke the combined emotions of admiration and fear which produce the readiness to conform in a truly authoritarian system. . . . Strong though often implicit emphasis is placed on punishment, which takes the form of bodily harm, social isolation, and/or shaming rather than loss of love [Levinson & Huffman, 1955, p. 257].

This hypothesized relationship between authoritarianism and sanctioning behavior has not yet been adequately tested using direct measures of sanctioning behavior; the existing evidence involves primarily self-report measures of sanctioning behavior.

It was primarily self-report data which led to the hypotheses represented in the items of the Adorno et al. (1950) Authoritarianism (F) Scale. Self-report data were also used by Hart (1957) in a study relating the F scores of mothers to the punishment techniques they used on their children. Hart found a moderate positive correlation between mothers' F scores and their reported use of "nonlove-oriented" punishment techniques, such as punishment by ridicule and physical punishment.

Although such self-report data deserve attention, it has recently become widely recognized that it is important to validate the F Scale by relating it to direct measures of relevant behavior. While the major interest has been in substituting "behavioral" criteria for fixed-alternative self-report instruments possibly laden with acquiescence set, it also seems worthwhile to substitute behavioral criteria for free-response self-report instruments.

The present study is an attempt to relate F to a direct measure of sanctioning behavior. The sanctions made available to subjects in the present study stem from the two sanctioning dimensions distinguished by Sears, Maccoby, and Levin (1957): positive versus negative, and love oriented versus object oriented. Thus the sanctions in the present study were positive love oriented ("You did well . . ."), positive object oriented (a monetary reward), negative love oriented ("You did poorly . . ."), negative object oriented (a monetary penalty), and no sanction. The sanctions studied by Hart (1957) most closely parallel the negative love-oriented and negative object-oriented alternatives.

On the basis of the above-quoted description of the sanctioning behavior of authoritarians, the following hypotheses were formulated:

1. The low-F individual will tend to use sanctions which are more love oriented than those used by the high-F individual.
2. The low-F individual will tend to use sanctions which are more positive than those used by the high-F individual.

METHOD

Subjects

One-hundred and twenty male students enrolled in introductory psychology at the University of Texas volunteered to undergo testing to fulfill a course requirement. The test included original F items which Byrne and Bounds (1964) found to correlate significantly with total F scores for a University of Texas sample. These included all the items of Form 45 of the F Scale, except items 22, 33, 38, 43, and 44; Items 66 and 67 from Form 78; and Item 41 from Form 60. Also included were five reversed F items, each of which was found by Byrne and Bounds (1964) to correlate significantly with total scores on the original F items listed above.

---

[1] This study was conducted under a grant from the University of Texas Research Institute.

The present investigators, after finding that F as measured by the original items correlated .43 with F as measured by the reversed items, added together each subject's z scores on the two variables. This procedure resulted, for each subject, in one F score which represented original and reversed F items equally and thus probably decreased the contaminating effects of acquiescence set.

The men in the lower and upper four-tenths of the resulting F distribution were then invited to volunteer for an experiment. The person who was to be the experimenter did not take part in this selection process, so that he would not be aware of subjects' F scores. Of the 20 low- and 22 high-F men who volunteered and were run, 2 high-F subjects were discarded for failure to follow instructions. The mean F scores on the original items for the final low- and high-F groups were 40.5 and 60.4 (possible range, 0–108), and the mean F scores on the reversed items for the same two groups were 4.4 and 8.9 (possible range, 0–20).

Measuring Sanctioning

The subjects arrived for the study by ones or twos, and each subject was seated in a separate cubicle. The experimenter then announced that the subject had previously answered items intended to measure "leadership ability," and had obtained a high leadership score. Now, in order to validate the leadership test, it was necessary to find out just how effective the subject really was in leading others. The subject was told that another subject, a "follower," was seated in an adjacent cubicle, and that his own job (although the follower did not know this) was to maximize the performance of the follower on a given task, using only a standard set of sanctions.

The follower's task, the leader was told, involved drawing X marks inside circles about ½ inch in diameter. An unseen buzzer would sound once to signal the beginning of a trial, and the "follower," as quickly as he could, would X in circles on a sheet containing 60 consecutively numbered circles. Twenty seconds later the buzzer would sound twice, signaling the end of that trial, and the follower would hand the performance sheet through a slot. After each trial, the leader's job was to enter the number of circles completed in a record sheet and decide what sanction he would use on that trial. The available sanctions were: poker chips representing \$.02, \$.01, \$−.01, and \$−.02; 3 × 5-inch cards reading "You did very well on the last trial"; "You did well on the last trial"; "You did poorly on the last trial"; "You did very poorly on the last trial"; and "Go on to the next trial." One and only one of these sanctions was to be delivered to the follower after each trial.

The poker chips handed through the slot to the follower would not be returned, but would be accumulated by the follower; subjects were told that at the end of the session, any net gains would actually be paid to the follower, and any net losses would actually be subtracted from the \$.50 the follower had been paid at the beginning of the experiment. The standard messages, which the leader would transcribe in his own handwriting before the trials began, would be returned each time by the follower after he had read them. This was to prevent them from accumulating like the monetary sanctions.

In reality, of course, it was the experimenter who sounded the buzzer, handed previously filled-in performance sheets through the slot, and received and recorded sanctions sent by the subject. Two subjects could be run simultaneously in adjacent cubicles by this method.

There were 15 trials, and the numbers of circles "completed" on trials 1–15 were 24, 26, 27, 29, 28, 33, 37, 42, 40, 40, 39, 37, 33, 28, and 33.

After the 15 trials, the subject was asked to evaluate the overall performance of the follower on a rating scale.

RESULTS AND DISCUSSION

In order to test the hypothesis that the low-F individuals would tend to use sanctions which were more love oriented than those used by the high-F individuals, a continuum was defined such that a love-oriented sanction (notes reading ". . . did very well . . . ," ". . . did well . . . ," ". . . did poorly . . . ," or ". . . did very poorly . . .") was scored 1, no sanction was scored 0, and an object-oriented sanction (\$.02, \$.01, \$−.01, \$−.02) was scored −1. This kind of scoring system, since it included all possible alternative sanctions, resulted in low- and high-F group means which both represented all trials equally.

While the low-F subjects did have a somewhat higher (more love-oriented) mean score on this variable than the high-F subjects, the difference was not significant.

It seems possible that the failure of this study to support the love-orientedness hypothesis may have been due to the love-oriented sanctions being less attractive than usual to the leaders. In both low- and high-F groups, the object-oriented sanctions were used almost twice as often as love-oriented ones. This may have been due to the absence in this study of leader motivation to conserve monetary resources and of the follower-leader solidarity which may be a necessary prerequisite for the effective use of love-oriented sanctions.

In order to test the hypothesis that the low-F individual will tend to use sanctions which are more positive than those used by the high-F individual, a continuum was defined such that a positive sanction (\$.01 or \$.02, or notes reading ". . . did well . . ." or ". . . did very well . . .") received a score of 1, no sanction

TABLE 1

USE OF VARIOUS SANCTIONS BY HIGH- AND LOW-F SUBJECTS

Ss	Mean frequency of the sanctions:		
	Positive	No	Negative
High F	7.90	1.25	5.85
Low F	8.85	2.25	3.90
t	1.08	2.06*	2.63*

* $p < .05$, $df = 38$, 2-tailed test.

received a score of 0, and a negative sanction (\$−.01 or \$−.02, or notes reading ". . . did poorly . . ." or ". . . did very poorly . . .") received a score of −1. The mean score for the low-F group on this continuum was 4.95, and that for the high-F group was 2.05. The difference between these two means was in the predicted direction and significant ($t = 1.86$, $df = 38$, $p < .05$) by one-tailed test.

If such a result were obtained in a field study where follower performance was not controlled, one possible explanation would be that the followers performed differently for the two types of leader. However, such an explanation is not possible here, since follower performance was exactly the same for low- and high-F leaders. Another conceivable explanation would be that low-F subjects used sanctions which were more positive than those used by high-F subjects because they evaluated the follower's performance more favorably. However, when the mean evaluations of overall follower performance by the two groups of subjects were compared, they were almost exactly equal: 4.65 for the low-F subjects and 4.70 for the high-F subjects, on a scale running from 1 (extremely poor) to 7 (excellent).

Thus it appears that low- and high-F subjects, even when faced with identical follower performances which they evaluate identically, react with different sanctioning behavior. How can this be explained? The instrumental nature of the sanctioning behavior in the present study brings to mind the possibility that the low- and high-F subjects differ in their beliefs concerning the effectiveness of different sanctions. Further research would be needed to test this possibility.

While the positivity hypothesis and its test involved treating positive, no, and negative sanctions all as values of one variable, it is also of interest to look at each of these sanctions separately. Table 1 shows that the low-F subjects used positive sanctions slightly more frequently than the highs, used no sanction significantly more frequently than the highs, and used negative sanctions significantly less frequently than the highs. This set of findings implies, of course, that the distribution of the positivity variable had a somewhat smaller variance for the low-F subjects than for the highs.

The directions of these differences in usage of positive and negative sanctions were expected. As for the use of no sanction, the authors had no expectations. However, the obtained difference does seem consistent with the implication of the passage quoted at the beginning of this paper to the effect that the authoritarian has especially great concern for controlling subordinates.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BYRNE, D., & BOUNDS, C. The reversal of F Scale items. *Psychological Reports,* 1964, **14**, 216.

HART, I. Maternal child-rearing practices and authoritarian ideology. *Journal of Abnormal and Social Psychology,* 1957, **55**, 232–237.

LEVINSON, D. J., & HUFFMAN, P. Traditional family ideology and its relation to personality. *Journal of Personality,* 1955, **23**, 251–273.

SEARS, R. R., MACCOBY, E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957

(Received June 9, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 225–228

RAPPORT, AWARENESS, AND VERBAL REINFORCEMENT OF ATTITUDE

CHESTER A. INSKO AND KENT W. BUTZINE

University of North Carolina

Ss contacted by telephone were reinforced with "good" for agree or disagree responses indicating a favorable or unfavorable attitude toward pay TV. The results indicated that reinforcement had a significant effect which interacted with the type of experimentally manipulated rapport existing between E and S. The reinforcement effect was greater with positive than with negative rapport. The correlations of the conditioning or attitude scores with each of Dulany's 3 types of awareness produced generally inconsistent results. In view of the fact that the experiment was conducted with Ss who did not know that they were in an experiment, it was concluded that the results could not be an artifact of the demand characteristics of an experimental situation.

Numerous successful attempts to verbally reinforce attitudes have been reported (Ekman, 1958; Hildum & Brown, 1956; Insko, 1965; Krasner, Knowles, & Ullman, 1965; Singer, 1961). In spite of the fact that these studies have used various issues in various contexts it would be unwise to suppose that verbal reinforcement will always be effective in all situations. One obvious variable possibly affecting the success of verbal reinforcement is the degree of rapport existing between the experimenter and subject. A balance or consistency approach implies that there is a greater tendency to agree with liked than disliked persons. Also, agreement with liked persons may be more hedonically satisfying than agreement with disliked persons. The present study is an attempt to investigate the role of rapport (liking-disliking of the experimenter) in a verbal reinforcement situation.

The genuineness of the verbal reinforcement phenomenon is currently being questioned because of some evidence indicating that successful conditioning only occurs with subjects who are aware of the purpose of the experimental manipulations (e.g., Levin, 1961). Dulany (1962) has distinguished between three types of awareness: reinforcement hypothesis (awareness of the reinforcement contingency), behavioral hypothesis (awareness of what the experimenter expects the subject to do), and behavioral intention (intention to do what the experimenter wants the subject to do). These types of awareness are each assessed with different postexperimental questions. Questions asking the subject if he was aware of when the experimenter said "good" assess reinforcement hypothesis; questions asking the subject if he was aware of what he was supposed to say or what the experimenter wanted him to say assess behavioral hypothesis; questions asking the subject how he went about responding and what he attempted to do assess behavioral intention. Dulany obtained evidence indicating that all three of these types of awareness correlate with conditioning but that the reinforcement hypothesis correlates least highly and behavioral intention correlates most highly. All three types of awareness thus are placed in a theoretical network which serves to predict the conditioning effect. Theoretically, it is behavioral intention which produces the reinforced response. Behavioral intention, however, is dependent upon behavioral hypothesis which is in turn dependent upon reinforcement hypothesis.

As Insko and Oakes (1966) have pointed out, this entire awareness question raises two separate issues. One issue relates to whether or not there is learning without awareness of the reinforcement contingency. This is an old problem that has been of historical concern in the field of learning. The second issue relates to whether or not awareness of the demand characteristics of the experimental situation (Orne, 1962) is a necessary prerequisite for the occurrence of conditioning. This latter issue relates to the genuineness or possible artifactuality of the conditioning phenomenon. The two issues roughly correspond to Dulany's distinction between reinforcement hypothesis and behavioral hypothesis, although Dulany does not make evident the possible artifactual implications of the latter as opposed to the former type of awareness. In any event, the major point relates to the fact that successful attempts to verbally reinforce attitudes in nonlaboratory situations (Hildum & Brown, 1956; Insko, 1965) indicate that the phenomenon is not an artifact of the laboratory. It still may be the case, of course, that some kind of awareness is correlated with or even causes conditioning, but this is no basis for doubting the genuineness of the phenomenon. The present study attempts to assess each of Dulany's three types of awareness in a nonlaboratory context.

METHOD

Subjects

Seventy-six undergraduates at the University of North Carolina were selected from the student directory, and an attempt was made to contact each by telephone. The original sample was random except for the fact that it included not more than one person on any given dormitory floor. Out of the original sample, 12 were not included in the final analysis. Eight had either withdrawn from school or moved to an unknown address, two refused to participate in the study, one was intoxicated, and one incorrectly claimed to be an acquaintance of the experimenter. The final sample of 64 subjects included an equal number of males and females.

Procedure

The general procedure involved asking subjects to "strongly agree," "agree," "disagree," or "strongly disagree" to each of a series of 12 opinion statements regarding pay TV. The statements are presentd in Table 1. Half of the subjects were reinforced with "good" for responses indicating a positive attitude toward pay TV, and half for responses indicating a negative attitude. Immediately following the conditioning procedure all subjects were asked the questions contained in Dulany's (1962) awareness questionnaire.

TABLE 1

PAY TV STATEMENTS

Viewer selectivity, not pay TV, is the way to improve TV quality.
Pay TV firms want to "horn in" on facilities that free TV has built.
Pay TV would end the current "wasteland" produced on free TV.[a]
A large segment of intelligent America is being virtually ignored by the producers of current free TV.[a]
Pay TV would bring live Broadway plays and exclusive sports events to everyone's television set.[a]
Many TV commercials are well executed and enjoyable to watch.
Pay TV is within the price range of most people owning sets.[a]
Pay TV will bring cultural events to outlying areas which otherwise would have no opportunity to see them.[a]
Pay TV would not offer much you cannot see now for free.
The reduction or elimination of commercials from television would hurt the American economy.
Since advertising either cannot, or believes it cannot, support certain high-quality programs, a ready-made audience awaits widespread pay TV.[a]
No private interest, such as those who propose to operate pay TV, has the right to profit from publicly owned airwaves by charging for receiving programs over those airwaves.

[a] Pro statements, all others being con.

Independent Variables

Three independent variables were manipulated in the experiment: sex of subject, direction of reinforcement (pro or con), and rapport (positive or negative). All of these variables involved between-subject comparisons. Direction of reinforcement was manipulated by reinforcing with "good" agree responses to pro statements and disagree responses to con statements or agree responses to con statements and disagree responses to pro statements. The statements were selected on the basis of of extensive pilot testing so that, for all 12 statements combined, an approximately equal number of pro and con responses were obtained.

Rapport was manipulated by having the experimenter mildly ingratiate or insult the subject for his responses to an initial series of two questions. The experimenter represented himself as a low-status high school student in order to magnify the effect of negative rapport. The verbatim account of the procedure is as follows:

> I'm a member of the Public Opinion Club of Chapel Hill High School. We're conducting a survey of University student attitudes toward television, and, in particular, toward pay-TV. This will take only one or two minutes of your time, and your opinions will be kept anonymous. Would you mind participating in this survey?
>
> First, I'd like you to give me the name of your best liked television show (or shows), aside from so-called special show.
>
> Positive rapport: That's a good show. I watch it a lot myself.
>
> Negative rapport: Mmmm . . . . [name of show in a derogatory tone] Humph.
>
> The next question is in the form of a multiple choice. Would you say that the effects of television on the development and maturity of children are, on the whole, A. very beneficial, B. somewhat beneficial, C. neutral, D. somewhat harmful, E. very harmful.
>
> Positive rapport: O.K. Fine. That's what most of the research indicates, by the way.
>
> Negative rapport: . . . did you say!!! ___? Mmmm . . . , well, anyway.
>
> Now I'm going to read to you a list of statements taken from newspapers and magazines, which express various opinions toward pay-TV. For each statement, I'd like you to indicate your reaction in one of four ways. With each statement, please tell me if you: 1. strongly agree, 2. agree, 3. disagree, or 4. strongly disagree.
>
> Do you have any questions? Then we will begin. Here is the first statement . . . .

The experimenter then read in turn each of the 12 statements contained in Table 1.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE ATTITUDE SCORES IN THE EIGHT EXPERIMENTAL CONDITIONS

	Males				Females			
	Positive rapport		Negative rapport		Positive rapport		Negative rapport	
	Pro	Con	Pro	Con	Pro	Con	Pro	Con
M	20.50	15.25	18.50	17.25	19.75	15.12	19.25	16.62
SD	2.51	2.96	2.27	3.15	2.32	4.09	3.06	1.99

Dependent Variables

Responses to the 12 opinion statements were scored 0, 1, 2, or 3 with the high score indicating a favorable attitude toward pay TV. Responses to the awareness questions were coded according to Dulany's rules so as to obtain an assessment of each of the three types of awareness. The percentage of perfect agreement between two judges was 90.63 for reinforcement hypothesis, 96.88 for behavioral hypothesis, and 98.44 for behavioral intention. The final obtained awareness scores were based on the mean ratings of the two judges.

RESULTS

The mean and standard deviation estimates of the attitudes scores are contained in Table 2 and an analysis of variance of these data in Table 3. Other than the total between effect the only significant effects are the expected ones for reinforcement and the Reinforcement × Rapport interaction. The main effect for reinforcement indicates that the direction of reinforcement had a significant effect upon attitude across all conditions, and the interaction signifies that reinforcement was more potent in the positive-rapport condition than in the negative-rapport condition. Within the negative-rapport condition the reinforcement effect, however, is still significant ($t = 2.09$, $p < .05$, $df = 30$). For all of the pro reinforcement conditions the mean is above 18 (the neutral point of the scale), and for all of the con reinforcement conditions the mean is below 18.

TABLE 3

ANALYSIS OF VARIANCE OF THE ATTITUDE SCORES

Source	*df*	*MS*	*F*
Rapport (A)	1	1.00	
Sex (B)	1	0.56	
Reinforcement (C)	1	189.06	23.09**
A × B	1	1.00	
A × C	1	36.00	4.39*
B × C	1	0.56	
A × B × C	1	3.99	
Total (between)	7	33.17	4.05**
Within	56	8.19	

* $p < .05$.
** $p < .01$.

TABLE 4

CORRELATIONS BETWEEN ATTITUDE AND AWARENESS IGNORING SEX

	Positive rapport		Negative rapport	
	Pro	Con	Pro	Con
Reinforcement hypothesis	+.238	−.318	+.453*	−.097
Behavioral hypothesis	+.410	+.217	+.603*	−.258

[a] 1-tailed.
* $p < .05$.

The within-cell correlations of attitude with reinforcement hypothesis and behavioral hypothesis are presented in Tables 4 and 5. Table 4 presents the correlations ignoring the sex classification, and Table 5 presents the correlations ignoring both the sex and rapport classifications. In view of the nonsignificance of the sex effect, the collapsing of this variable, at least, is justified. Attitude was not correlated with behavioral intention because of the lack of variability in the awareness scores. All but two of the subjects were scored at the neutral point of the scale. Subjects typically responded to the behavioral intention question by indicating either that they had no intention or that they merely intended to state their true opinion regarding each statement. The majority of the correlations of attitude within reinforcement hypothesis and behavioral hypothesis are not significant. Because of the way the variables were scored, the correlations in the pro conditions theoretically should have been positive and the correlations in the con conditions should have been negative. Except for the correlation of .217 relating attitude and behavioral hypothesis in the positive-rapport–con-reinforcement condition this expectation is upheld. This one correlation, however, should perhaps not be taken too seriously because of the lack of variability in awareness in that particular cell.

DISCUSSION

The correlational data possibly point to some weak relationship between awareness and condi-

TABLE 5

CORRELATIONS BETWEEN ATTITUDE AND AWARENESS IGNORING SEX AND RAPPORT

	Pro	Con
Reinforcement hypothesis	+.256	−.215
Behavioral hypothesis	+.414*	−.007

* $p < .05$.

tioning, although the details of Dulany's (1962) theoretical network are certainly not supported. According to Dulany, behavioral intention should have correlated most highly with attitude. As a matter of fact, however, so little variability was found in the assessment of behavioral intention that there was no point in computing the correlations. They would have all been at or close to zero. This may reflect the fact that the subjects were not aware of their behavioral intentions or that they would not admit being aware of them. On the basis of the existing data it is impossible to say. In general, the awareness data do not appear very enlightening.

The data support the conclusion that verbal reinforcement can change attitudes in a nonlaboratory situation and furthermore that the reinforcement effect, to some extent, depends on the type of rapport existing between experimenter and subject. The sizable main effect for reinforcement and the significant reinforcement effect within the negative-rapport condition both testify to the general robustness of the phenomenon.

Robustness of the reinforcement phenomenon in a nonlaboratory situation is not the sort of thing that one would have expected on the basis of previously obtained awareness correlations. The subjects in the present experiment did not know that they were in an experiment and still showed the predicted effect. If demanded characteristics are to account for these results, then the meaning of the term will have to be severely redefined.

REFERENCES

Dulany, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. *Journal of Personality,* 1962, **30** (Suppl.), 102–109.

Ekman, P. A comparison of verbal and nonverbal behavior as reinforcing stimuli of opinion responses. Unpublished doctoral dissertation, Adelphi College, 1958.

Hildum, D. C., & Brown, R. W. Verbal reinforcement and interviewer bias. *Journal of Abnormal and Social Psychology,* 1956, **53**, 108–111.

Insko, C. A. Verbal reinforcement of attitude. *Journal of Personality and Social Psychology,* 1965, **2**, 621–623.

Insko, C. A., & Oakes, W. F. Awareness and the "conditioning" of attitudes. *Journal of Personality and Social Psychology,* 1966, **4**, 487–496.

Krasner, L., Knowles, J. B., & Ullman, L. P. Effect of verbal conditioning of attitudes on subsequent motor performance. *Journal of Personality and Social Psychology,* 1965, **1**, 407–412.

Levin, S. M. The effect of awareness on verbal conditioning. *Journal of Experimental Psychology,* 1961, **61**, 67–75.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17**, 776–783.

Singer, R. D. Verbal conditioning and generalization of prodemocratic responses. *Journal of Abnormal and Social Psychology,* 1961, **63**, 43–46.

(Received June 13, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 228–232

NEED APPROVAL, DEFENSIVE DENIAL, AND DIRECTION OF AGGRESSION IN A FAILURE-FRUSTRATION SITUATION[1]

LEROY H. FORD, JR. AND MICHEL HERSEN[2]
State University of New York at Buffalo

The defensive denial interpretation of social desirability (SD) responding was tested. 80 male undergraduates were failed in an achievement-oriented situation. A postexperimental questionnaire yielded experimentally independent measures of extrapunitiveness, intropunitiveness, and impunitiveness, with SD controlled. It was predicted that high-SD Ss would manifest denial by being less intropunitive than lows and would supplement this defense by being more extrapunitive. High scorers on a Lie type of SD scale were less intropunitive than lows, while high Edwards SD Ss were more extrapunitive than lows. These results support the defensive denial interpretation of the Lie type of SD scale, but a question is raised as to whether a pure need-approval hypothesis could account for these results.

In recent years, there has been an increasing interest in the motivational determinants of social desirability (SD) responding. Since the development of the MMPI Lie and *K* scales, the tendency to put oneself in a favorable light

---

[1] Portions of this paper were presented at the Rocky Mountain Psychological Association Convention, Albuquerque, New Mexico, May, 1966.

[2] Now at Psychotherapy Associates of Fairfield, Fairfield, Connecticut.

on personality questionnaires has been viewed as reflecting a general denying or repressing mode of defense (Dahlstrom & Welsh, 1960). More recently, an interpretation of the Lie type of SD scale in terms of a need for social approval has been advanced by Crowne and Marlowe (1960). However, as research on the approval-motivation construct has accumulated, the original conception has been revised to include defensiveness as a component of the need for approval (Crowne & Marlowe, 1964).

The purpose of the present study was to shed further light on the defensiveness interpretation of SD by investigating its relationship to Rosenzweig's (1945) three major "direction of aggression" variables, intropunitiveness, extrapunitiveness, and impunitiveness, following failure on an achievement-oriented task. In previous studies relating SD-test rseponses to nontest behavior, little attention has been given to whether the nontest behavior itself might simply be another instance of SD responding. In the present study the effects of SD in the measurement of direction of aggression were controlled.

A prediction regarding intropunitiveness seemed to be most directly derivable from the defensiveness interpretation. If SD responding represents a threat-motivated denial of personal faults and inadequacies, it should manifest itself in a tendency to avoid placing the blame for failure on oneself. Accordingly, it was hypothesized that, following personal failure, high-SD subjects would manifest less intropunitiveness than lows.

With regard to extrapunitiveness, it was thought that the high-SD subject might attempt to strengthen the effect of denial by externalizing or projecting the blame. Accordingly, it was hypothesized that, following failure, high-SD subjects would manifest greater extrapunitiveness than lows.

With regard to impunitiveness, no clear hypothesis seemed to be derivable. The impunitive reaction did not seem to have any clear potential for bolstering the high-SD person's denial of his own faults and inadequacies. This response represents an attempt to ignore the frustrating and aggression-arousing implications of the situation (Rosenzweig, 1945). Presumably, its defensive function is to minimize the person's aversive emotional reactions to the situation. Although denial presumably serves a similar purpose, the forms of the two defenses did not seem to be sufficiently similar to permit a prediction about impunitiveness on the basis of the defensive-denial construct.

Two SD scales were employed in the present study, the Ford Social Desirability (FSD) Scale (Ford, 1964a) and the Edwards Social Desirability (*SD*) scale (Edwards, 1957). The FSD Scale, like the Marlowe-Crowne Social Desirability (M-C *SD*) scale (Crowne & Marlowe, 1960), is based on the Lie scale model, but makes use of a forced-choice format to eliminate acquiescence bias. The FSD Scale correlates in the .70s and .80s with the M-C *SD* scale and, except for a small difference in acquiescence bias, the two scales are probably functionally unitary.

The predictions of the present study were made primarily with the FSD Scale in mind and, by extension, the M-C *SD* scale. However, at least one line of evidence indicates that the Edwards *SD* scale, which is not a Lie type of scale, may also reflect defensive denial (Ford, 1964b); on the basis of this evidence, it was also thought that the predictions would hold for the Edwards *SD* scale.

METHOD

Construction of Postexperimental Questionnaire

Seven questions directly related to the failure-frustration situation to be experienced by the experimental subjects were constructed by the authors. For each of the seven questions, three responses, one each to represent the extrapunitive, intropunitive, and impunitive dimensions, were constructed. For example, one question was, "How do you feel about the results on this test?" The three responses pertaining to this question were, "I don't think the results mean very much"; "I should have achieved a higher score"; and "I usually enjoy taking tests, irrespective of results." These represented the extrapunitive, intropunitive, and impunitive dimensions, respectively. Thus, the questionnaire contained 21 items, 7 for each of the categories of direction of aggression.[3] There was 100% agreement among the authors and two independent judges[4] as to the direction of aggression represented by each of the items.

The questionnaire was presented to 24 male undergraduate psychology majors for SD judgments. The judges received a SD rating booklet with the following instructions:

> A University of Buffalo student is scheduled to participate in an experiment but when he arrives he is given a paper-and-pencil group intelligence test. The experimenter hands out the test and instructs the subjects that this is a timed test and that in a similar group at the University of Rochester, the average student was able to complete the test successfully within the allotted time. When the signal is given that time is up, the student is still working on the test and is not close to being finished. Immediately following this procedure, each student is asked to complete a

---

[3] A copy of the complete questionnaire may be obtained from the first author.

[4] The authors thank Richard Eisler and Arnold Markowitz for making these judgments.

questionnaire relating to his experience during testing.

> The questionnaire given to the students appears on the following pages. After each question you will find a list of three possible answers that a person might give to that question. You are asked to read each question and to rate the social desirability of each answer to the question. That is, for each answer, indicate how desirable or undesirable it would be considered, from the point of view of society, for a person to give that answer to the question, after experiencing the situation described above.

The judges indicated their ratings on a 5-point scale, with a rating of 1 indicating "strongly undesirable" and a rating of 5 indicating "strongly desirable." The mean ratings were 3.07, 3.07, and 2.81 for the extrapunitive, intropunitive, and impunitive subscales, respectively. A second set of SD ratings was obtained from a group of 19 male introductory psychology students using a 9-point rating scale and slightly modified instructions which emphasized somewhat more the failure aspect of the frustration situation. The mean ratings were 5.1, 5.5, and 4.5 for extrapunitive, intropunitive, and impunitive, respectively. Only the difference between the intropunitive and impunitive means in the second set of ratings was significant ($p < .05$ by Wilcoxon matched-pairs signed-ranks test). It is important to note that all of the means cluster around the neutral point (3.00 for the first set and 5.00 for the second set). The neutral SD values of the three response categories indicate that SD should be a minor determinant of extrapunitive, intropunitive, and impunitive scores (Edwards, 1957).

Subjects

A sample of 80 male subjects was selected at random from the introductory psychology subject pool at the State University of New York at Buffalo. Half of the subjects were assigned to the experimental group and half to the control group. As part of the course requirement, all subjects had previously completed the FSD Scale and the Edwards SD scale. For both scales, high- and low-SD subjects were divided at the median. A comparison of mean SD scores of experimentals versus controls for the high- and low-FSD and Edwards scale groups revealed no significant differences and indicated that the corresponding experimental and control groups were equated for SD scores on both scales.

Procedure

The experiment was conducted in two group sessions, one for the experimental subjects and one for the control subjects. All subjects were given a 180-item symbol-substititution test form. The following instructions were given by the experimenter:

> The purpose of this experiment is to obtain normative data for a new test of intelligence. This is a timed test of how quickly you can learn new associations. In order for the study to be meaningful, it is important that you do the best you can. We have recently tested a similar group of undergraduates at the University of Rochester and the average student was able to complete at least 160 of the 180 items. You can perhaps see that the person who does well is able to learn the associations most quickly. So, as you are taking the test try to learn as many associations as you can.

In both groups, the experimenter gave the signal to start. Control-group subjects were permitted sufficient time so that all completed the entire 180 items. In the experimental group, at a predetermined point the experimenter announced that the time limit was up. The time allowed was sufficiently short that no subject was able to complete more than about 140 items. Thus, the experimental subjects were subjected to the failure experience of not being able to complete the test and of falling short of the reported performance of the average University of Rochester student.

After completing the symbol-substitution test, all subjects filled out the postexperimental questionnaire. The subjects responded to each of the 21 items on a 5-point rating scale, with a rating of 1 indicating that the particular response was "very unlike" the subject's own reaction, and a rating of 5 indicating that it was "very much like" the subject's own reaction. Independent scores were obtained for the extrapunitive, intropunitive, and impunitive subscales. Thus, the measurement here is "normative" rather than "ipsative," and the difficulties inherent in ipsative measurement are avoided. This procedure differs from the standard measurement of direction of aggression (Rosenzweig, 1945) which is essentially ipsative.

RESULTS

The effect of the experimental manipulation of failure on each of the three direction of aggression variables was determined by *t*-test comparisons between the experimental and control groups. The experimental group was significantly more intropunitive than the control group ($t = 3.48$, $df = 78$, $p < .001$), but the differences in extrapunitiveness and impunitiveness did not approach significance. Thus, the overall effect of the experimental treatment was to elicit intropunitive feelings, and there was no overall tendency for the experimental subjects to show either an extrapunitive or impunitive reaction.

The mean intropunitive, extrapunitive, and impunitive scores of the high-SD and low-SD experimental and control groups are presented in Table 1. For each of the three dependent variables, the differential effect of the experimental treatment on the high- and low-SD groups was determined by computing the significance of the

TABLE 1

MEAN INTROPUNITIVE, EXTRAPUNITIVE, AND IMPUNITIVE SCORES FOR HIGH- AND LOW-SD GROUPS

Group	Intropunitive		Extrapunitive		Impunitive	
	M	*SD*	*M*	*SD*	*M*	*SD*
High FSD						
Experimental	21.35	3.05	20.30	2.98	18.95	3.61
Control	19.90	3.99	20.45	3.50	20.45	2.91
Low FSD						
Experimental	24.00	4.76	22.20	4.82	20.40	4.44
Control	19.10	4.36	20.75	4.81	19.50	3.20
High Edwards *SD* scale						
Experimental	23.35	4.28	22.60	3.72	20.30	4.80
Control	20.30	3.54	20.30	4.23	21.45	2.84
Low Edwards *SD* scale						
Experimental	22.00	4.04	19.90	4.04	19.05	3.17
Control	18.70	4.64	20.90	4.17	18.50	2.56

Note.—$N = 20$ in each cell.

difference between the differences for independent groups as described by Scott and Wertheimer (1962, p. 266). For the intropunitive variable, a significant differential effect was found for the FSD Scale but not for the Edwards *SD* scale; for the extrapunitive variable, there was a significant effect for the Edwards *SD* scale, but not for the FSD Scale; and for the impunitive variable, none of the differential effects approached significance.

For the intropunitive variable, there was a significant difference between the experimental-control difference shown by the high-FSD subjects and the difference shown by the low-FSD subjects ($t = 1.88$, $df = 76$, $p < .05$, one-tailed). A further analysis of the intropunitiveness data showed that the high-FSD experimental subjects did not differ significantly from their controls ($t = 1.26$), whereas the low-FSD experimental subjects were significantly more intropunitive than their corresponding control subjects ($t = 3.31$, $df = 38$, $p < .005$, one-tailed). Also, in the experimental condition the high-FSD subjects were significantly less intropunitive than the lows ($t = 2.04$, $df = 38$, $p < .05$, one-tailed), while in the control condition, the high and low groups did not differ significantly ($t = .59$).

For the extrapunitive variable, there was a significant difference between the experimental-control difference shown by the high-Edwards *SD* groups and that shown by the low-Edwards *SD* groups ($t = 1.82$, $df = 76$, $p < .05$, one-tailed). A further analysis showed that the high-Edwards *SD* experimental subjects were more extrapunitive than their controls ($t = 1.78$, $df = 38$, $p < .05$, one-tailed), whereas the low-Edwards *SD* experimental subjects did not differ significantly from their controls ($t = .75$). Also, within the experimental condition, the high-Edwards *SD* subjects were more extrapunitive than the lows ($t = 2.14$, $df = 38$, $p < .05$, one-tailed), while in the control condition, the high- and low-Edwards *SD* scale groups did not differ significantly ($t = .44$).

DISCUSSION

The finding that the low-FSD subjects manifested an intropunitive reaction to failure-frustration while the high-FSD subjects did not confirms the first prediction derived from the defensive denial hypothesis. This finding indicates that the subject who scores high on the Lie type of scale defends against feelings of failure by being less self-castigating and less willing to acknowledge his own responsibility for his failure. Since, as indicated by the judges' ratings, the intropunitiveness scale was quite neutral with respect to SD, an interpretation of these results in terms of a generalized SD response set is not tenable.

An important question to raise in considering the intropunitiveness findings has to do with their relevance to a "pure" need-approval interpretation of the Lie type of SD scale. According to this interpretation, high-SD persons respond in socially sanctioned ways in order to gain the approval of others (Crowne & Marlowe, 1960). The predictions to be made on the basis of this hypothesis seem relatively straightforward when the behavior in question has clear SD connotations. But the hypothesis does not seem to offer a clear basis for prediction when the SD connotations of the behavior are not salient. Since SD was controlled in the present study, it appears that the results cannot readily be explained on the basis of a simple need-approval hypothesis. However, as we have noted, Crowne & Marlowe (1964) have recently expanded their original interpretation to include defensiveness as a "component" of the need for social approval; with this adjustment, the present results can be accounted for.

Making such an adjustment does not, of course, rule out the possibility that the Lie type of SD scale might be more fruitfully and parsimoniously viewed as primarily a measure of defensive denial. This interpretation has had wide acceptance among MMPI users for many years; many of the recent need-approval studies (Crowne & Marlowe, 1964) have made recourse to a defensiveness interpretation in order to account for their results. If the defensive denial construct has more predictive and explanatory power than the "pure" need-approval construct, then the currently popular interpretation in terms of need for approval may need to be revised. It has been suggested elsewhere (Ford, 1964a, 1964b) that the concept of need for approval implies an approach tendency, while the concept of defensiveness implies an avoidance tendency. Although these tendencies may simply represent opposite ends of a single behavioral continuum, they may instead be at least partially independent response dispositions. As with other approach and avoidance tendencies, they may develop on the basis of different antecedent conditions and, therefore, may arise from potentially independent life experiences. In view of these considerations, it would seem to be important to give further attention to the clarification of these issues.

The expectation that the high-FSD subject might bolster his defensive denial by externalizing the blame for his poor performance was not supported; the high-FSD subjects did not show more extrapunitiveness than the lows. However, this result may not necessarily be inconsistent with the defensive denial interpretation. While denial is a repressive defense, externalization or "projection" is an expressive or sensitizing type of defense. The present results are consistent with previous findings that repressive defenses are characteristic of high scorers on the Lie type of scale. In addition, however, they suggest that in the present situation, where the use of experimentally independent measures permits the manifestation of both repressive and expressive modes of defense, high scorers make use of the former, but not the latter.

With regard to the Edwards *SD* scale, the results are not consistent with a previous suggestion (Ford, 1964b) that Edwards *SD* scale scores may be viewed as a reflection of defensive denial. However, that suggestion was based on a relatively narrow rather than diversified set of data, since the behaviors from which defensiveness was inferred in these studies all involved expectancy statements of some kind. Thus, the defensiveness hypothesis must remain a most tentative one, as far as the Edwards *SD* scale goes, and it is quite possible that further research will suggest the need for reinterpretation. This possibility is further suggested by the finding that, in the experimental condition, the high-Edwards *SD* scale subjects were more extrapunitive than the lows. Since the high-Edwards *SD* subjects, following failure-frustration, manifest greater externalization than the lows, it appears that the highs are characterized by an expressive rather than repressive mode of defense.

REFERENCES

CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology,* 1960, 24, 349-354.

CROWNE, D. P., & MARLOWE, D. *The approval motive.* New York: Wiley, 1964.

DAHLSTROM, W. F., & WELSH, G. S. *An MMPI handbook.* Minneapolis: University of Minnesota Press, 1960.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

FORD, L. H., JR. A forced-choice, acquiescence-free, social desirability (defensiveness) scale. *Journal of Consulting Psychology,* 1964, 28, 475. (a)

FORD, L. H., JR. Social desirability, defensive denial, and expectancy for success. Paper read at Midwestern Psychological Association, St. Louis, May 1964. (b)

ROSENZWEIG, S. The picture association method and its application in a study of reactions to frustration. *Journal of Personality,* 1945, 14, 3-23.

SCOTT, W. A., & WERTHEIMER, M. *Introduction to psychological research.* New York: Wiley, 1962.

(Received June 13, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 233–236

THE ANGERED:
THEIR SUSCEPTIBILITY TO VARIETIES OF HUMOR[1]

EARL S. DWORKIN AND JAY S. EFRAN
University of Rochester

Relationships between anger and humor were studied by playing tape recordings to angered Ss. They listened to hostile humor (N = 10), nonhostile humor (N = 10), or documentary and musical recordings (N = 10). A separate control group of 20 Ss listened to the hostile (N = 10) or nonhostile (N = 10) humor without having been angered. Humor significantly reduced reported feelings of anger and anxiety (as assessed by the Nowlis-Green Mood Adjective Check List). Angered Ss appreciated hostile humor more than did nonangered Ss.

The small body of experimental literature which investigates the relationship between motivational states and humor preferences has produced largely negative findings (Byrne, 1958; O'Connell, 1960; Sears, Hovland, & Miller, 1940; Singer, 1965). However, with the exception of the study by Singer, facets of the design of these investigations limit the confidence that can be placed in their conclusions. For example, they do not independently measure the effects of arousal; therefore, negative findings may be due to a failure to adequately arouse the desired affect. Also, these studies use written jokes or cartoons. If these forms of humor are less potent than material presented orally by a live performer, the results obtained thus far may not hold for all humor. Finally, past research has focused primarily on whether people in a given state of arousal respond more positively to humor, but the effect of humor on mood has not been given sufficient attention.

This study investigates the relationship between anger and humor using (*a*) an independent index of arousal, (*b*) orally presented humor stimuli, and (*c*) a measure of the effect of humor on aroused anger. It is addressed to the following questions: What is the effect of humor on anger? What effect does anger have on a person's appreciation of humor? Does the content of the humor make a difference? These questions lead to the formulation of the following more specific hypotheses: (*a*) Subjects who feel angry rate humor as "funnier" than do nonangered subjects; (*b*) feelings of anger are mitigated by exposure to humor, and this effect is maximized when the content of the humor is hostile; (*c*) angered subjects appreciate hostile jokes more than nonhostile jokes, whereas nonangered subjects appreciate nonhostile jokes more than hostile jokes; (*d*) there is a positive relationship between the appreciation of humor and its effectiveness in decreasing anger.

METHOD

Subjects

The subjects were 50 male undergraduate students from the University of Rochester and the Rochester Institute of Technology.

Stimulus Materials

Six hostile and six nonhostile humor stimuli were chosen from a pool of 75 taped selections from various comedians' record albums.[2] The level of hostility of each selection had been judged by eight advanced clinical psychology graduate students on a

---

[1] This report is a revised and expanded version of a paper read at the Eastern Psychological Association Convention, New York, April, 1966.

[2] The six nonhostile humor selections used were taken from the following recordings: *The Best of Mike Nichols and Elaine May* (Mercury Records MG 20997, Side 1, Band 4); *Bill Cosby is a Very Funny Fellow, Right!* (Warner Brothers Records W 1518, Side 2, Band 2); *Golden Hits of the Smothers Brothers* (Volume 2, Mercury Records MG 21089, Side 1, Band 3); *The Button Down Mind of Bob Newhart* (Warner Brothers Records W 1379, Side 2, Band 3); *The Best of the Stan Freberg Shows* (Capital Records WBO-1035, Side 1, Band 2; *The First Nine Months are the Hardest* (Capital Records T 2034(Side 1, Band 2).

Only five of the six hostile humor selections are available, since one of the selections came from a promotion record which has since been discontinued. The five used were taken from the following records: *An Evening with Mike Nichols and Elaine May* (Mercury Records OCM 2200, Side 1, Band 1); *Woody Allen* (Colpix Records CP 518, random selection from Side 1 since this record is a continuous band); *Bob Newhart Faces Bob Newhart* (Warner Brothers Records W 1517, Side 1, Band 2); *Mike Nichols and Elaine May Examine Doctors* (Mercury Records 20680, Side 1, Band 5); *Bill Cosby is a Very Funny Fellow, Right!* (Warner Brothers Records W 1518, Side 2, Band 3).

TABLE 1

LENGTH OF TAPED SELECTIONS AND THEIR MEAN "FUNNINESS" AND "HOSTILITY" RATINGS

Stimuli	N	Funniness M^a	Hostility M^b	Time (in min.)
Hostile humor tapes	6	4.08	4.32	21
Nonhostile humor tapes	6	4.06	1.58	21
Control tapes	6			20

[a] Based on the ratings of 30 undergraduate students using a 7-point scale.

[b] Based on the judgments of eight advanced clinical psychology graduate students using a 5-point scale.

5-point scale, ranging from 1 ("Humor depends solely on its general cleverness, incongruity, surprise, silliness, funny sounds, etc. There is no element of sarcasm or belittlement towards persons and/or institutions") to 5 ("Humor involves very sarcastic and vicious swipes at people and/or institutions. They are made to look foolish and stupid"). The overall reliability for these ratings, as assessed by a concordance coefficient (Winer, 1962), was .91. The hostile ($M = 4.32$) and nonhostile ($M = 1.58$) selections were matched for both "funniness" (having been previously rated by a group of 30 undergraduates on a 7-point "funniness" scale) and time (each tape lasting approximately 20 minutes). Six control tapes, consisting of documentary readings and music, were also used. The overall data pertaining to the stimulus material is presented in Table 1.

Arousal Measure

A modified form of the Nowlis-Green Mood Adjective Check List (MACL, Nowlis, 1965) was used to assess the subject's affective state. It consisted of 28 words concerning mood. Subjects were instructed to indicate how they felt at a given moment by responding to each word on a 4-point continuum from *no* (definitely do not feel this way at the moment) to *vv* (definitely feel this way at the moment). Two cluster scores were derived from this measure, a hostility score, consisting of the sum of the responses to seven words (angry, fed up, defiant, antagonistic, hostile, rebellious, annoyed), and an anxiety score, consisting of the ratings for three words (clutched up, fearful, jittery).

Procedure

A group of 30 subjects were treated in the following manner by a businesslike experimenter. He told them to write a short autobiographical sketch indicating what they were like when they were younger, what they are like now, and what they would like to be when they grew up. When they finished, he read what they had written and said,

> No, no—what'd you say your name was? Well, I really don't have much use for this. In the first place, the writing is so sloppy I can hardly read it. In the second place, the grammar is about eighth-grade level. And besides, it's much too long. I didn't expect you to run off at the mouth like this! Did I ask you for a life history? You'd better do this: just print me three sentences—one about your past, one about your present, and one about your future. And make sure you print it this time so I can read it.

When they had finished, the experimenter read the sentences and mumbled, "Well, this will have to do." Following this, he handed the subject a copy of the MACL and said, "Now I want you to fill out this form. The instructions on it are very simple. If you read it carefully, do you think you'll be able to understand it?" Immediately after this anger-arousal procedure, 20 of these experimental subjects were given a humor rating form, were told they would be required to rate the "funniness" of some taped humor selections, and were left alone to listen to the tapes and make their ratings.[3]

TABLE 2

t TESTS OF SIGNIFICANCE BETWEEN THE PRE- AND POSTHOSTILITY SCORES FOR THE THREE ANGERED GROUPS

Groups	M prescore^a	SD	M postscore	SD	M difference	t
Hostile humor ($N = 10$)	7.20	5.31	2.50	2.59	4.70	3.43**
Nonhostile humor ($N = 10$)	6.70	4.16	4.00	3.61	2.70	3.69**
Control tapes ($N = 10$)	7.00	3.05	4.80	3.22	2.20	1.81

[a] Comparison between mean prescores yields no significant differences.

** $p < .01$.

---

[3] The humor rating scale was a 7-step scale ranging from 1 ("not funny at all"), through 4 ("moderately funny"), to 7 ("extremely funny"). The instructions to the subjects were:

> You will be listening to six taped selections by several comedians. We would like you to judge how funny you think each selection is on a 7 point scale from 1 (not funny at all) to 7 (extremely funny). For example, if you thought selection 1 was somewhat more than moderately funny, but not in the extremely funny category, you would place a check in box number 5. There

TABLE 3

t TESTS OF SIGNIFICANCE BETWEEN THE PRE- AND POSTANXIETY SCORES FOR THE THREE ANGERED GROUPS

Groups	*M* prescore[a]	*SD*	*M* postscore	*SD*	*M* difference	*t*
Hostile humor ($N = 10$)	3.00	2.53	0.75	1.62	2.25	3.64**
Nonhostile humor ($N = 10$)	2.20	2.08	0.70	1.89	1.50	2.37*
Control tapes ($N = 10$)	2.10	2.00	1.00	0.86	1.10	1.57

[a] Comparison of mean prescores yields no differences.
* $p < .05$.
** $p < .01$.

Ten of these subjects listened to the hostile selections and 10 to the nonhostile selections. After they listened to the tapes, the experimenter returned and administered a second MACL. The remaining 10 angered subjects listened to the control tapes and were instructed to rate them on "interest level." A postexperimental interview revealed that 3 subjects had correctly guessed the purpose of the experiment; their data were not used. A separate control group of 20 subjects listened to the hostile ($N = 10$) or nonhostile ($N = 10$) humor without having been angered by the experimenter.

RESULTS

A comparison between the prehumor MACL hostility cluster scores for the 30 experimental subjects and the 20 control subjects clearly indicated that hostility was aroused in the experimental groups ($t = 7.07$, $p < .001$). Thus, the precondition that the experimental subjects be angry was met.

Table 2 presents the data concerning the MACL hostility scores for the angered subjects before and after exposure to the tapes. Both the hostile and nonhostile humor resulted in significant decreases in hostility scores ($p < .01$), where the control tapes did not. However, the hostile humor did not cause a significantly greater drop in hostility scores than the nonhostile humor ($t = 1.29$, *ns*).

A similar analysis was performed on the anxiety cluster scores of the MACL. The results paralleled the data for hostility; that is, both hostile and nonhostile humor resulted in significantly decreased anxiety, whereas the control tapes did not (see Table 3).

To test the effect of arousal on humor preferences, an analysis of variance was performed (see Table 4). Results indicated that the main effect was not significant, but that there was a significant interaction between anger and humor content ($F = 4.93$, $p < .05$). Further analysis of the mean "funniness" ratings (Table 5) indicated that this interaction was primarily due to the fact that hostile humor was judged to be significantly funnier by angered ($M = 3.81$) than by nonangered ($M = 3.17$) subjects ($t = 2.33$, $p < .05$). The mean nonhostile humor rating by angered subjects was 3.36, and for nonangered subjects it was 3.58.

Rank-order correlations were computed between subjects' humor ratings and their hostility-change scores. The correlation between hostile humor judgments and hostility-change scores was .04, and between judgments of nonhostile humor and hostility-change scores, .34. Neither of these was statistically significant.

DISCUSSION

While other studies have failed to show a relationship between drive arousal and humor preferences, the results of the present study indicate that humor preferences are sensitive to at least one kind of arousal—anger. However, anger does not make people more appreciative of humor in

TABLE 4

ANALYSIS OF VARIANCE FOR HUMOR JUDGMENTS

Source	*SS*	*MS*	*F*
Arousal (A)	0.44	0.44	1.19
Type of humor (B)	0.01	0.01	<1
A × B	1.82	1.82	4.93*
Error	13.38	0.369	
Total	15.65		

* $p < .05$, $df = 1/36$.

TABLE 5

MEAN "FUNNINESS" RATINGS FOR HUMOR SELECTION

Ss	Hostile humor	Nonhostile humor
Angered ($N = 20$)	3.81	3.36
Nonangered ($N = 20$)	3.17	3.58

---

will be a 10–20 second break between selections for you to mark your choice.

general; rather, it appears to lead the individual to respond selectively to humor of hostile content. In other words, people respond more strongly to humor stimuli which in some way parallel or relate to their current cognitions.

In terms of the effect that exposure to humor has on feelings of hostility and anxiety, the results indicate that humor significantly mitigates such feelings. This study does not, of course, shed light on the mechanism by which this mitigation occurs.

The results of our study do not allow us to predict individual differences in anger reduction from humor-appreciation scores. Our failure to find a significant correlation between humor ratings and subsequent feelings of anger is not congruent with a finding reported by Singer (1965). With $N = 6$, he found a negative correlation of .987 between humor appreciation and subsequent level of aggression. If this finding is replicable, a number of factors might account for the discrepancy between our results and his. In his study, Negro subjects listened to taped speeches by a white segregationist. They then listened to a Negro comedian talk pointedly about segregationists. Thus, both the source of the anger arousal and the content of the humor were very specifically directed at a single issue. Individual reaction patterns to this specific topic might well produce differences in response both to the humor and to the arousal technique. In our study, with much less specific content, individual differences were either not impressive, or, if present, were masked by the insensitivity of our measuring instruments.

To summarize, the present study has attempted to explore the effect that humor has on anger, and the effect that feelings of anger have on humor appreciation. Anger seems not to have a general effect on humor preferences. Rather, there is an interaction between anger and the content of the humor. Specifically, people who are angry respond more positively to hostile humor than do people who are not angry. Also, exposure to humor seems to decrease reported feelings of anger and anxiety.

REFERENCES

BYRNE, D. Drive level, response to humor, and the cartoon sequence effect. *Psychological Reports*, 1958, 4, 439–442.

NOWLIS, V. Research with the mood adjective check list. In S. Tompkins & C. Izard (Eds.), *Affect, cognition, and personality*. New York: Springer, 1965. Pp. 352–389.

O'CONNELL, W. The adaptive functions of humor. *Journal of Abnormal and Social Psychology*, 1960, **61**, 263–270.

SEARS, R., HOVLAND, C., & MILLER, N. Minor studies of aggression: Measurement of aggression behavior. *Journal of Psychology*, 1940, 9, 275–279.

SINGER, D. *The carthartic function of humor*. Paper read at Eastern Psychological Association, New York, April 1965.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received June 15, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 237-240

PUNISHMENT AND AGGRESSION IN FANTASY RESPONSES OF BOYS WITH ANTISOCIAL CHARACTER TRAITS[1]

JUDITH B. SCHAEFER AND MARTIN NORMAN

Newton-Baker Project, Newtonville, Massachusetts

Fantasy responses of 2 age groups (preadolescent, N = 14; adolescent, N = 20) of antisocial and matched normal boys were compared for aggression and punishment variables. Control adolescents gave significanty more aggression themes than antisocial adolescents (p = .05). Preadolescents did not differ for aggression themes. External punishment themes more often followed aggression in the stories of antisocial than control Ss, regardless of age (p's = .01, .07). The incidence of internal punishment themes was low for all groups. However, control adolescents gave more internal punishment themes than preadolescent controls (p = .05). These results support the view that punishment variables are relevant to the relationship between overt antisocial behavior and fantasy aggression.

Results of a few studies suggest that overt aggressive behavior may be more consistently correlated with punishment variables in fantasy or with ratios of punishment to aggression themes than to aggression themes only (Lesser, 1958; Purcell, 1956). The present study was designed to explore these relationships by comparing punishment and aggression in fantasy responses of boys with antisocial character traits[2] and a control group of normal boys. Previous studies have lacked either a normal group or a group with clearly diagnosed antisocial tendencies.

Purcell (1956), viewing antisocial behavior as an impulse-control disequilibrium, compared aggression and external and internal punishment themes in the TAT responses of an antisocial aggressive and a nonaggressive group of Army draftees who were psychiatric referrals. In addition to more aggression themes in the responses of the antisocial group, he found more external and internal punishment themes in the nonaggressive group. The internal punishment difference was highly significant; the external punishment difference was significant only for punishment themes unrelated to aggression.

Studies of younger subjects have not clearly distinguished internal and external punishment. Using a scale that incorporated both variables, Shore, Massimo, and Mack (1964) reported increased internal punishment or guilt in TATs of delinquent adolescents who showed behavioral improvement in a vocational treatment program. Mussen and Naylor (1954), focusing primarily on external punishment, presented suggestive evidence that lower-class boys who feared punishment were not as overtly aggressive as less fearful subjects. However, Lesser (1958) found both aggression themes and punishment by authority figures positively correlated with overt aggression.

Differences in the ages and antisocial tendencies of the subjects may underlie some of the inconsistent findings reported above. Only Purcell's (1956) subjects were late adolescent or adult, and only Lesser's (1958) subjects were drawn from a presumably normal population. Moreover, the projectives, as well as specific measures, differed among the studies. Nevertheless, punishment variables appear relevant to the relationship between overt and fantasy aggression, suggesting the major hypothesis of this study: the relationship between aggression and punishment will differ in fantasy responses of antisocial boys and presumably normal boys. More specifically, the combined results of earlier studies suggest that antisocial boys, compared to normals, will tell (*a*) more stories with aggression themes, (*b*) more stories with aggression themes followed by external punishment, but (*c*) fewer stories with internal punishment.

METHOD

Subjects

Subjects ($N = 34$) were selected from the public schools of Newton, Massachusetts, on the basis of questionnaires and interviews with teachers who, in effect, rated them on antisocial behavior. Seventeen

---

[1] This research was supported by Research Grant 2-R-11-MH11-2 from the National Institute of Mental Health.

The authors wish to thank Paule Daum and Hyman Kempler for their assistance.

[2] The boys had been referred to the Newton-Baker Project, a 5-year research and demonstration effort concerned with the treatment of predelinquent boys and their families, sponsored by the Judge Baker Guidance Center in cooperation with the Newton Public Schools.

boys (the antisocial group) were among a larger group accepted at the Newton-Baker Project after additional psychiatric interviewing suggested they had antisocial character defects. The remaining 17 subjects (the nonreferred group) were boys who did not exhibit antisocial behavior in school, but who were similar in age and socioeconomic status to the boys in the antisocial group.

Ages ranged from 11 years, 1 month to 18 years, 1 month for the antisocial group and from 11 years, 5 months to 18 years, 1 month for the nonreferred group at the time of testing. The age difference of matched pairs ranged from 0 to 10 months with a mean difference of 4.4 months. Two fathers in each group were professional types; all other fathers were skilled or white-collar workers. For the antisocial group, IQ (Wechsler Intelligence Scale for Children; Wechsler-Bellevue Intelligence Scale, Form II; or Wechsler Adult Intelligence Scale) scores ranged from 80 to 120 with a mean of 103. Scores ranged from 83 to 131 for the nonreferred group with a mean of 111.

Procedure

Each subject was individually tested with a battery of nine Murray TAT cards, seven cards selected from a similar projective devised by Seymour Epstein, and one card from the Michigan Picture Test. The test included TAT cards 1, 2, 3BM, 4, 5, 10, 12M, 13B, 17BM; Epstein cards EM1 (boy sitting on a dock facing a body of water), EM3 (boy approaching a maternal figure standing in a doorway, a bus in the background), EM4 (boy with a scowl on his face walking with fists clenched), EM8 (two boys sitting, one crying, the other with his left arm around the shoulder of the crying boy), EM16 (boy with a gun standing over a bleeding man, a man running in the background), EM17 (the back of a man in judge's robes facing a boy with a policeman on one side, a crying woman and a man on the other), EMA (man shaking his finger at a boy); and MP6 (four boys and two girls leaning over a checker game).

The 578 thematic protocols were scored by a graduate psychology student for aggression and external and internal punishment. Although criteria for identifying aggression were similar to those used in earlier studies, the range of aggressive acts that appear in fantasy suggested that antisocial aggression should not be confounded with less serious types of aggression. Thus, the presence of two categories of aggression was separately recorded: antisocial and impulsive. It seemed more desirable to score for presence rather than frequency of aggression because the stories of antisocial subjects tended to be shorter than those of nonreferred subjects. Consequently, only antisocial aggression (the more extreme behavior) was scored when both types occurred in a single story. External and internal punishment were scored only in stories with an aggression (antisocial or impulsive) theme.

Specifically, antisocial aggression was defined as any behavior, implicit or explicit, aimed at hurting another physically, or behavior typically regarded as antisocial, for example, killing, fighting, attacking, stealing, destroying property, truancy. Impulsive aggression was defined as any behavior less antisocial than above and without intent to hurt another physically, for example, quarreling, verbally abusing, defying or resisting authority (staying out too late, sneaking out to avoid violin practice), doing something wrong (unspecified), accidental and relatively minor destruction of property. External punishment was defined as any behavior on the part of an authority figure aimed at punishing, restricting, or getting retribution for an aggressive or impulsive act of another, for example, sending a boy to reform school, restricting him to his room, spanking, or verbally chastising him. Internal punishment was defined as feelings of regret or remorse, feeling sorry for an aggressive act on the part of the person who committed the act.

For each subject, antisocial aggression, external punishment, and internal punishment scores were obtained by counting the stories in which the appropriate response occurred. A total aggression score was also determined by adding the impulsive aggression score to the antisocial aggression score. For any given subject, the total aggression score could not exceed 17 (the number of pictures presented), and punishment scores could not exceed the total aggression score.

Independent scoring of 170 stories by three psychologists showed 84–90% agreement for total aggression themes, 87–89% for antisocial aggression themes, 86–97% agreement for external punishment, and 86–89% agreement for internal punishment.

The Mann-Whitney U test (Siegel, 1956) was used to compare the groups for themes of aggression and punishment. Corrections for ties were made when appropriate.

RESULTS

Total Group Differences

Table 1 presents mean scores of the thematic variables for the total 17 antisocial and 17 nonreferred subjects, for adolescent and preadolescent subgroups, and the appropriate Mann-Whitney U and p values. All tests are two-tailed.

Contrary to expectation, the nonreferred group told more stories with aggression themes than the antisocial group, a difference, however, that was not statistically significant. As predicted, the antisocial group told relatively more stories with external punishment. Since the external punishment score is dependent upon the aggression score, the ratio of the external punishment scores to the total aggression score was ranked. These ratios were greater for the antisocial group. Conversely, the frequency of unpunished aggressive acts (the total aggression score minus the external punishment score) was significantly greater for the nonreferred subjects.

TABLE 1

MEAN SCORES FOR FANTASY MEASURES, MANN-WHITNEY U, AND p VALUES FOR ANTISOCIAL AND NONREFERRED BOYS

Fantasy measure	Adolescent				Preadolescent				Total			
	Antisocial ($N = 10$)	Nonreferred ($N = 10$)	U	p	Antisocial ($N = 7$)	Nonreferred ($N = 7$)	U	p	Antisocial ($N = 17$)	Nonreferred ($N = 17$)	U	p
Total aggression	6.5	8.3	25.5	.05	7.14	6.71		*ns*	6.64	7.65		*ns*
External punishment	2.6	2.5		*ns*	3.86	2.28	13	.12	3.12	2.41		*ns*
Unpunished aggression[a]	3.9	5.8	17	.01	3.29	4.43	17.5	*ns*	3.41	5.23	78	.02
External punishment/total aggression	0.40	0.30	36.5	*ns*	0.55	0.33	10	.07	0.47	0.32	92.5	.07
Internal punishment	0.70	2.10	30.5	.12	1.29	0.14	16.5	*ns*	0.94	1.29		*ns*

[a] Total aggression score minus external punishment score.

Age Differences

Because inspection of the raw scores suggested that younger subjects differed from older subjects with respect to some of the TAT variables, the data were analyzed separately for subjects between 14 and 18 years old (adolescent) and for subjects between 11 and 13 years (preadolescent).

As indicated in Table 1, nonreferred adolescents told significantly more stories with aggression themes than antisocial adolescents. They also told more stories with internal punishment. However, antisocial adolescents told relatively more stories with external punishment, as indicated by significantly fewer stories with unpunished aggressive acts.

For the preadolescent groups ratios of external punishment to total aggression were greater for antisocial than nonreferred subjects, and the difference in absolute frequency of external punishment approached statistical significance. In contrast to the older groups, the antisocial subjects told more stories with aggression themes than did nonreferred subjects, but this trend was not statistically significant. Contrary to expectation, the mean number of internal punishment themes was greater for antisocial than nonreferred preadolescents; however, the responses of one antisocial subject were largely responsible for this difference.

Because mean scores for aggression and punishment varied somewhat with age, preadolescent and adolescent groups within the diagnostic groups were compared for differences in aggression and punishment theme frequencies using the Fisher test (Siegel, 1956). Only one significant difference in 10 comparisons was found: more nonreferred adolescent subjects, as compared to preadolescent nonreferred subjects, gave internal punishment responses following themes of aggression ($p = .05$).

DISCUSSION

The results supported the major hypothesis of this study: the relationship between aggression and punishment in fantasy responses of antisocial boys differed from that of nonreferred boys.

However, not all of the secondary hypotheses were supported. Antisocial subjects did not tell more stories with aggression themes. On the contrary, nonreferred adolescent boys in this study told more stories with aggression themes than did antisocial adolescents. Although several factors could account for the differences between this finding and earlier studies reporting a positive correlation between fantasy and overt aggression, the age variable seems most relevant. Subjects in the previous studies were either older or younger than the adolescent group in this study. Purcell's (1956) subjects were between 17 and 25 and Lesser's (1958) were under 14. Although the age range for Mussen and Naylor's (1954) subjects was from 9 to 15, the number of boys who corresponded in age to the older subjects in the present study cannot be determined from their report.

Given the nature of adolescence, the greater frequency of aggression themes in the fantasy of adolescent, as contrasted to preadolescent, nonreferred boys is not surprising. However, the fact that a similar age difference did not exist for the antisocial groups (in fact, the mean score for antisocial adolescents is smaller than the mean score for antisocial preadolescents) suggests the operation of defense mechanisms, such as denial, which are less marked in healthier adolescents.

As predicted, antisocial subjects, regardless of age, gave more external punishment themes following an aggression theme. This is consistent with Lesser's (1958) finding that punishment by authority is associated with overt aggression. It is not consistent with Purcell's (1956) results or the suggestion of Mussen and Naylor (1954) that overt aggression and fear of punishment are inversely related. Again, the differences in subjects and specific measures used in the various studies may underlie the lack of agreement. The age factor, in particular, may account for the difference between these results and Purcell's (1956) study of older subjects, whereas degree of antisocial tendency may underlie the inconsistency with Mussen and Naylor (1954). All of the subjects in the latter study were referred to the Bureau of Juvenile Research in Columbus, Ohio, for behavior which had brought them into conflict with police and school authority. Comparisons were made between the most and least overtly aggressive.

Purcell (1956, p. 14) has suggested that antisocial subjects tend to justify their aggressive reactions by pointing to the harsh treatment imposed on them by others. Thus, retaliation or counteraggression, as external punishment was defined here, reinforces defensiveness and reduces guilt over antisocial behavior. Although the results of this study do not prove such speculations, the fact that antisocial boys, more often than nonreferred boys, told stories in which aggression is followed by external punishment lends support to Purcell's view.

Internal punishment, a measure inversely related to overt aggression in subjects 16 years or older (Purcell, 1956; Shore et al., 1964), is infrequent in responses of subjects less than 14 years. The greater frequency of internal punishment in stories of adolescent, as contrasted to preadolescent, nonreferred subjects suggests a developmental change not evident among antisocial boys. Since internal punishment has been equated to guilt, its appearance in juxtaposition with fantasy aggression after age 14 raises several provocative theoretical and empirical questions.

In general, the results of this study indicate that fantasy responses of antisocial boys differ from nonreferred boys in the relative frequency with which external punishment follows aggression. In addition, the fantasy of adolescent, but not preadolescent, antisocial boys tends to have fewer aggression themes. Comparisons between and within the two diagnostic groups at two ages suggest differential change with age in the frequency of aggression and punishment themes of these groups, pointing up the importance of age or developmental stage in the relation of overt aggressive behavior and aggression and punishment variables in fantasy.

REFERENCES

LESSER, G. S. Conflict analysis of fantasy aggression. *Journal of Personality,* 1958, **26,** 29–41.

MUSSEN, P. H., & NAYLOR, H. K. The relationships between overt and fantasy aggression. *Journal of Abnormal and Social Psychology,* 1954, **49,** 235–240.

PURCELL, K. The TAT and antisocial behavior. *Journal of Consulting Psychology,* 1956, **20,** 449–456.

SHORE, M. F., MASSIMO, J. L., & MACK, R. The relationship between levels of guilt in thematic stories and unsocialized behavior. *Journal of Projective Techniques,* 1964, **28,** 346–349.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received June 16, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 2, 241–244

ATTITUDE DIFFERENTIATION AND EVALUATIVE SCALES OF THE SEMANTIC DIFFERENTIAL[1]

S. S. KOMORITA AND ALAN R. BASS
Wayne State University

On the assumption that evaluative scales of the semantic differential might be fruitfully differentiated into a no. of dimensions, 220 undergraduate students were administered 16 evaluative, bipolar scales and asked to evaluate 2 issues: (a) American foreign policy in Vietnam, and (b) draft deferments for married students. The intercorrelations between the 16 scales were factor analyzed for each issue and 3 distinct factors were obtained on each issue: (a) a functional-utilitarian factor, (b) an affective-emotional factor, and (c) a moral-ethical factor. The 1st factor supports the utilitarian function of attitudes proposed by Katz and the 3rd factor suggests a normative basis for attitudes. It is concluded that evaluative scales of the semantic differential do not always yield a general, evaluative dimension and the utility of differentiating such scales is discussed.

In recent years, an increasing number of investigators have been employing the semantic differential (Osgood, Suci, & Tannenbaum, 1957) as a measure of attitude (e.g., Triandis & Fishbein, 1963). Probably the main reason for its use is that it is convenient and readily available without the necessity of scale construction—one only has to select a set of bipolar scales with high loadings on the evaluative factor. Moreover, the evaluative scales of the semantic differential have been shown to correlate highly with Thurstone attitude scales and have been shown to have relatively high reliability.

In their original series of studies, Osgood et al. pointed out one of the problems of using the semantic differential as a measure of attitude, namely, the problem of concept-scale interaction. They described the difficulty they had in finding a set of scales which loaded heavily on a certain factor across a variety of concepts. In a study by Shaw (cited in Osgood et al., 1957, p. 176; see also Triandis, 1960), it was found that certain concepts seemed to evoke or arouse an evaluative reaction more than others. Hence, the meaning of certain evaluative scales seemed to change as a function of the particular concept being judged. The main implication of Shaw's results is that one cannot rely completely on the factor loadings reported by Osgood et al., but must also select a set of scales which is appropriate for the concept to be evaluated. This problem is more serious for the measurement of attitudes since Osgood et al. reported data showing that evaluative scales are less stable and more susceptible to variations across concepts than nonevaluative scales.

The fairly typical Scale × Concept interactions found by Osgood and others in studies using evaluative-type semantic differential scales suggest that the affective component of attitudes is not necessarily unidimensional, but may be differentiated depending upon the particular attitude domain and subjects being studied. It is of interest to note that Chein (1951), in describing prejudice and discrimination, had classified affective orientations into two types: net positive and negative affect, and specific affect. Under specific affects he listed such dimensions as distrust-trust, contempt-admiration, and undifferentiated-differentiated emotion. Many attitude theorists, however, when referring to attitude differentiation, think in terms of the differentiation of the cognitive component. According to Katz (1960), for example, differentiation refers to "the number of beliefs or cognitive items contained in the attitude, and the general assumption is that the simpler the attitude in cognitive structure the easier it is to change [p. 168]." Krech, Crutchfield, and Ballachey (1962), on the other hand, proposed the term *multiplexity* to refer to "the number and variety of the elements or parts making up a component [p. 142]." The major implication of this concept is that multiplexity may refer to the differentiation of the affective and behavioral components as well as the cognitive component, whereas for Katz, differentiation refers primarily to the cognitive component.

The primary purpose of the present study is to amplify and illustrate the problem of concept-scale interaction when using the semantic differential as a measure of attitude, and to show how such interactions may be related to attitude differentiation. As indicated previously, Osgood et al. (1957) were well aware of this problem, and the authors' objective is primarily to caution investigators of the dangers of assuming comparability of evaluative scores across different concepts (attitude objects, issues, persons), as well as to caution against the assumption of a unidimensional evaluative component even for a single concept.

---

[1] This study was supported, in part, by a grant from the Carnegie Foundation to the Center for the Study of Cognitive Processes, Wayne State University.

METHOD

Subjects. The subjects for this study were 220 undergraduate students enrolled in introductory psychology classes, and participation in this study fulfilled part of their requirements for the course.

Procedure. A set of 16 evaluative, bipolar scales were employed. Eleven of the scales were selected from scales reported by Osgood et al. to have high loadings on the evaluative factor. The remaining 5 scales were drawn from Roget's Thesaurus on the basis of similarity in meaning with the other 11 scales which were: attractive-unattractive, satisfactory-unsatisfactory, trustworthy-untrustworthy, agreeable-disagreeable, and approve-disapprove.

The 16 scales were group administered to subjects in classes of about 25 to 40 students. Using the conventional instructions recommended by Osgood et al., all groups were asked to evaluate two issues: American foreign policy in Vietnam, and draft deferments for married men. All subjects were given the Vietnam issue first and the draft issue next.

For all subjects combined, the intercorrelations between the scales were computed for each issue, and the correlation matrices were factor analyzed using a principal components solution and a varimax rotation.

TABLE 1

FACTOR MATRICES OF EVALUATIVE SCALES OF SEMANTIC DIFFERENTIAL[a]

Factor	Vietnam			Draft		
	I	II	III	I	II	III
approve	*.81*	.11	.36	*.80*	.24	.11
wise	*.83*	.06	.29	*.80*	.32	.18
valuable	*.84*	.01	.17	*.64*	.26	.37
kind	.27	.37	.68	.31	.67	.06
clean	.03	.41	*.66*	.16	.46	*.66*
beneficial	*.74*	.08	.33	*.67*	.23	.29
honest	.38	−.02	*.74*	.40	.09	.75
fair	.42	.07	.71	.77	.06	.37
good	.65	.17	.51	.72	.40	.20
nice	.21	.48	.58	.25	.76	.33
pleasant	−.01	*.80*	.14	.14	*.88*	.13
trustworthy	.47	.07	*.52*	.21	.10	*.85*
attractive	.18	*.81*	.05	.23	*.78*	.15
agreeable	.63	.53	.08	.57	.56	.16
satisfactory	*.69*	.22	.27	*.74*	.20	.33
sincere	.42	−.10	*.62*	.29	.19	*.70*
% total variance	29.75	14.00	22.90	29.12	21.68	18.43

[a] Highest and relatively pure loadings on each factor, common to the two issues, are italicized.

TABLE 2

COMPOSITE SCORE INTERCORRELATIONS OF EVALUATIVE SCALES

	Vietnam			Draft		
	I	II	III	I	II	III
Vietnam						
I	—	.23	.65	.02	−.07	.06
II		—	.20	.05	−.04	.03
III			—	.11	.11	.20
Draft						
I				—	.49	.62
II					—	.36
III						—

RESULTS

The rotated-factor matrices for the two issues are shown in Table 1, where it can be seen that three relatively distinct factors are present for each issue.[2] The scales with the highest and relatively pure loadings on each factor, *common to the two issues,* are italicized and may be summarized as follows: (*a*) Factor I: approve, wise, valuable, beneficial, statisfactory (also good and agreeable, but these are not as pure), (*b*) Factor II: pleasant, and attractive (also agreeable and nice, but not as pure), and (*c*) Factor III: clean, honest, trustworthy, and sincere (also fair, but not as pure).

Factor I may be interpreted as a "functional-nonfunctional" evaluative factor, and corresponds to what Katz and Stotland (1959) have described as the utilitarian function of attitudes; Factor II represents an "affective-emotional" factor and seems to be a noncognitive, evaluative factor; Factor III seems to represent a "moral-ethical" factor, and might be described in terms of the normative function of attitudes.

In order to determine the degree of factor independence, composite scores for each subject on each factor were obtained and intercorrelated. The following scales were used to define the composite scores: (*a*) Factor I: approve, wise, and valuable, (*b*) Factor II: pleasant and attractive, and (*c*) Factor III: honest, trustworthy, and sincere. Table 2 shows the intercorrelations between the composite scores for each issue as well as between the two issues. It can be seen that Factors II and III are relatively independent of each other on both issues, but Factors I and III are moderately correlated with each other. Moreover, the composite scores across issues also seem to be independent of each other and provide evidence for the generality of the three-factor structure.

---

[2] The correlation matrices may be obtained by writing to the authors.

Table 1 also demonstrates the problem of concept-scale interaction. This can be seen by the discrepancies in factor loadings between the two issues for the scales, "kind-cruel," "fair-unfair," and "nice-awful." All three of these scales are standard evaluative scales and have been shown by Osgood et al. (1957) to load highly and consistently on the evaluative dimension. To some extent, even the primary evaluative scale, "good-bad," demonstrates that the meaning of a scale may vary with the concept being evaluated. Although "good-bad" loads heavily on Factor I for both issues, on the Vietnam issue it also loads heavily (.51) on the moral-ethical dimension, while on the draft issue it loads moderately (.40) on the affective-emotional dimension.

DISCUSSION

The major implication of these results is that it would be highly misleading to assume that evaluative scales of the semantic differential will always result in a general, unitary evaluation of an attitude object or issue. These results indicate that such evaluative scales, when used as a measure of attitude, might be fruitfully differentiated into additional evaluative dimensions. Interestingly enough, Osgood et al. described an earlier study in which they found clusters of evaluative scales—one cluster they called "morally evaluative" (clean, tasty, sacred, fair, valuable), and another cluster they called "emotionally evaluative" (soft, calm, peaceful, relaxed, smooth, and bland). Hence the differentiation of the evaluative factor found in this study, for two of the dimensions at least, are not simply chance findings, and it is plausible that such an evaluative factor structure might occur quite frequently. Moreover, Chein (1951), in his classification of prejudice, included "policy orientations" as well as affective reactions. Policy orientations, according to Chein, consist of a personal commitment regarding judgments about "what should—or must—I do," and closely parallel the moral-ethical factor found in this study. Thus, all three of the factors found in this study were described and classified by Chein. Therefore, there seems to be some evidence and support, not only empirically, but conceptually as well, for the differentiation of the affective component of attitude.

With regard to the specific factors found in this study, the main implication of Factor I is that it provides partial support for the functional basis of attitudes proposed by Katz and Stotland (1959), and the notion of perceived instrumentality proposed by Peak (1955) and Rosenberg (1956). Clearly, the first factor corresponds closely to the utilitarian and instrumental motives underlying attitudes. Although the second and third factors, on the basis of the factor-score analysis, are somewhat correlated with the first, the interesting finding is that the utilitarian-instrumental factor is the dominant one, and accounts for the largest proportion of the variance.

An important methodological implication of these results is that they might provide, with further refinement, a convenient method of representing an individual's attitude structure. Instead of representing an individual's attitude by a single evaluative score, one might represent the person's attitude by a set of factor scores. Theories of inconsistency such as balance, dissonance, and congruity should recognize that individuals differ in tolerance for inconsistency, and it is reasonable to expect large individual differences in the degree to which people differentiate between the three dimensions of evaluation. A cognitively complex person, for example, may admit to himself that a given attitude object or issue is positively instrumental and attractive but quite unethical and immoral. The cognitively simple person, on the other hand, might avoid differentiating between the three factors or may even de-differentiate if the inconsistency was severe and could not be resolved.

As an example illustrating the utility of this approach to attitude measurement, consider the moral-ethical factor (normative function) found in this study. The presence of this kind of factor raises an interesting question concerning the effects of conflict of motives underlying a given attitude. Many of the subjects in this study, for example, approved of United States policy in Vietnam, thought it was wise, valuable, beneficial and satisfactory; however, some of them also felt it was very unattractive and unpleasant, or relatively untrustworthy, insincere, and dishonest. It is plausible, and perhaps quite likely, that such individuals, given such inconsistencies, would be least resistant to attitude change. The practical implication of such a situation for attitude change is quite clear.

The main problem with this approach is that the factor structure may change as a function of the particular attitude domain, and instead of a concept-scale interaction, we might be confronted with a factor-concept interaction. It seems fruitful, nevertheless, to differentiate between the three types of evaluative scales and to take account of this multidimensionality if the data exhibit such properties. Moreover, although some attitude domains may yield a general evaluative factor, for some subclasses of individuals there may be different degrees of differentiation between the factors, and one could derive discrepancies between factor scores for these subclasses of people. With this approach one could, for example, test whether cognitive complexity as measured by the magnitude of such difference scores is a generalized phenomenon or is relatively specific to certain attitude domains.

It would be of further interest to determine the type of attitude domain for which one is likely to find a distinct three-factor structure. Some persons, as attitude objects, might evoke a moral-ethical reaction more than others. The issues used in this study, however, referred to policies, rules, or regulations, and it is plausible that a moral-ethical factor is likely to occur when such policies or norms are perceived to be unfair and unethical for some segment of the population. Another important consideration is whether the two motivational bases of the attitude, utilitarian versus moral-ethical, are in conflict with each other. It is reasonable to assume that conflict between these two types of reactions is a necessary condition for a two- or three-factor structure. The fact that the factor-score intercorrelations for the Vietnam issue were smaller than for the draft issue lends support to this conjecture. Based upon this assumption, therefore, it could be concluded that there was more attitudinal conflict regarding the Vietnam issue than for the draft issue.

REFERENCES

CHEIN, I. Notes on a framework for the measurement of discrimination and prejudice. In M. Jahoda, M. Deutsch, & S. W. Cook (Eds.), *Research methods in social relations. Part I: Basic processes.* New York: Dryden Press, 1951. Pp. 381–390.

KATZ, D. The functional approach to the study of attitudes. *Public Opinion Quarterly*, 1960, 24, 163–204.

KATZ, D., & STOTLAND, E. A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 3. Formulations of a person and the social context. New York: McGraw-Hill, 1959. Pp. 423–475.

KRECH, D., CRUTCHFIELD, R. S., & BALLACHEY, E. L. *Individual in society.* New York: McGraw-Hill, 1962.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

PEAK, H. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1955.* Lincoln: University of Nebraska Press, 1955.

ROSENBERG, M. J. Cognitive structure and attitudinal effect. *Journal of Abnormal and Social Psychology*, 1956, **53**, 367–372.

TRIANDIS, H. C. A comparative factorial analysis of job semantic structures of managers and workers. *Journal of Applied Psychology*, 1960, **44**, 297–302.

TRIANDIS, H. C., & FISHBEIN, M. Cognitive interaction in person perception. *Journal of Abnormal and Social Psychology*, 1963, **67**, 446–453.

(Received June 16, 1966)

Volume 6 July 1967 Number 3

Journal of Personality and Social Psychology

Editor: Daniel Katz, *University of Michigan*

Associate Editor: Robert B. Zajonc, *University of Michigan*

Robert P. Abelson
Yale University
Joseph B. Adelson
University of Michigan
Gordon W. Allport
Harvard University
Norman H. Anderson
University of California, San Diego
Elliot Aronson
University of Texas
Albert Bandura
Stanford University
Leonard Berkowitz
Cornell University
James Bieri
University of Texas
Jack W. Brehm
Duke University
Donald R. Brown
University of Michigan
Eugene Burnstein
University of Michigan
Arnold H. Buss
Rutgers University
Launor F. Carter
Systems Development Corporation
Rosalind D. Cartwright
University of Illinois
Medical Center, Chicago
Dwight W. Chapman, Jr.
Vassar College
Irvin L. Child
Yale University
Nickolas B. Cottrell
Kent State University
William Dember
University of Cincinnati
Morton Deutsch
Columbia University
Seymour Feshbach
University of Colorado
Donald W. Fiske
University of Chicago
Basil S. Georgopoulos
University of Michigan
Edwin P. Hollander*
Tavistock Institute of Human Relations
London, England
Murray Horwitz
New York University
Herbert C. Kelman
University of Michigan
Daniel J. Levinson
Connecticut Mental Health Center, New Haven
Richard A. Littman
University of Oregon
Robert B. MacLeod
Cornell University
Melvin Manis
Veterans Administration Hospital
Ann Arbor, Michigan
David C. McClelland
Harvard University
Charles G. McClintock
University of California, Santa Barbara
William J. McGuire
Columbia University
Vincent Nowlis
University of Rochester
Harold L. Raush
University of Michigan
Milton E. Rosenbaum
University of Iowa
William A. Scott
University of Colorado
Janet Taylor Spence
University of Texas
Percy H. Tannenbaum
University of Wisconsin
Harry C. Triandis
University of Illinois
Joseph Veroff
University of Michigan

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The *Journal of Personality and Social Psychology* is devoted to basic research and theory in the broad field of social interaction and its determinants and effects. More specifically, it is concerned with the subareas of (*a*) social interaction and group process, (*b*) interpersonal perception and attitude change, (*c*) the psychological aspects of formal social systems and less structured collective phenomena (e.g., the public opinion process and mass movements), (*d*) the psychological analysis of cultural forms and products (e.g., language), (*e*) the socialization process both at the child and adult levels, (*f*) social motivation and personality dynamics, (*g*) the structure of personality, and (*h*) the relation of personality to group process and social systems. Animal studies as they relate to an understanding of social processes are definitely appropriate for the journal. Methodological articles and studies of the development of personality instruments are not excluded but are considered marginal unless presented as part of a substantive piece. Where existing journals already serve a given specialized interest within the broad domain of the field, articles appropriate to this area of specialization will have low priority for the *Journal of Personality and Social Psychology*.

The JOURNAL is issued monthly, three volumes per year. The subscription price is $30.00 annually (foreign, $30.50). Single issues are $3.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor-Elect:

Dr. William McGuire
Journal of Personality and Social Psychology
604 Mathematics Hall
Columbia University
New York, New York 10027

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the Editor or from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR
Managing Editor

JOANNE BERGEN
Technical Editor

HENRY S. SALMON
Advertising and Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.
Copyright © 1967 by the American Psychological Association, Inc.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 245–254

DECISION FREEDOM AS A DETERMINANT OF THE ROLE OF INCENTIVE MAGNITUDE IN ATTITUDE CHANGE[1]

DARWYN E. LINDER, JOEL COOPER, AND EDWARD E. JONES

Duke University

In the forced-compliance paradigm, attitude change following a counterattitudinal performance has been shown to be both a direct (reinforcement prediction) and an inverse (dissonance prediction) function of the amount of incentive offered. An experiment successfully demonstrated that if S feels free not to comply attitude change will be inversely related to incentive magnitude, but that the positive relationship will hold if this freedom is reduced. It was hypothesized that the procedure of an earlier study by Rosenberg, whose results supported the reinforcement prediction, inadvertently reduced Ss' freedom not to comply. When this procedure was closely replicated in a 2nd experiment, the positive relationship was again found, but when the procedure was modified to make a decision not to comply a viable alternative for S, the inverse relationship resulted. The 2 experiments together show that a low incentive arouses dissonance, leading to attitude change, only when the person remains free to decide against compliance after he has been fully informed about the incentive. If the incentive is announced after the person is committed to compliance, a reinforcement effect obtains.

If a person can be induced to behave publicly in a manner that does not follow from his private attitudes, he will experience cognitive dissonance. The magnitude of dissonance will be greater when there are few reasons for complying than when there are many reasons (Festinger, 1957). This dissonance may be reduced by an accommodating change in private attitude if other ways of reducing dissonance are not available. Thus a person who has been induced to behave in a counterattitudinal fashion will change his private attitude more the less he has been rewarded for complying. Festinger and Carlsmith (1959) found support for this proposition in a study in which subjects were persuaded (for $1 or $20) to extol the attractiveness of a dull and tedious task for the benefit of the next subject. Also, Cohen (1962) found that Yale students who were induced to write essays in favor of the New Haven police later showed more positive attitudes the smaller the incentive they had been offered to write the essay.

Rosenberg (1965) has recently questioned the generality of this proposed relationship and has suggested that subjects in the Festinger and Carlsmith (1959) and the Cohen (1962) experiments must have considered the incentive excessive and, because of the "evaluation apprehension" that subjects in psychology experiments commonly feel, those in the high-incentive conditions may have interpreted the experiment as one testing their honesty and autonomy. To resist influence in the face of a substantial bribe, therefore, would cause the experimenter to evaluate them favorably. Alternatively, Rosenberg suggests that the subjects may have suspected deception in the high-incentive condition and angrily resisted confirming the perceived hypothesis. Either reaction might conceivably account for the obtained inverse relationship between incentive amount and degree of ultimate congruence between attitude and behavior.

Rosenberg proceeded to conduct an experiment loosely replicating Cohen's (1962), the major difference being that one experimenter provided the incentive for essay writing and another measured the subsequent attitude. The two tasks were presented to the subjects as unrelated, and thus there was presumably no chance for evaluation apprehension to affect the results. Rosenberg found that attitude and behavior were most congruent when subjects were offered $5 for writing essays and least in line when they were offered $.50. Rosenberg's results are clearly at variance with the apparent dissonance prediction and with the findings obtained by Cohen and by Festinger and Carlsmith. Instead, the results seem to favor a reinforcement position or, as Rosenberg would prefer, a theoretical position that considers the effects of reinforcement in the context of an affect-cognition consistency model. In Rosenberg's view, either the expectation or the receipt of reward strengthens and stabilizes the cognitions associated with the counterattitudinal statement—the greater the reward, the greater the stabilizing effect. There is then a change in attitudinal affect in the interests of cognitive-affective consistency.

---

[1] This experiment was facilitated by National Science Foundation Grant 8857. We are indebted to H. B. Gerard for his valuable suggestions.

Nuttin (1964) carefuly replicated Rosenberg's experiment and found that—even with evaluation apprehension removed in the same manner—inferred attitude change varied inversely with the amount of incentive offered. While Nuttin's results were of only borderline significance, they clearly offered more support for the dissonance proposition than for the counterproposition reflecting reinforcement theory. Aronson (1966) has criticized Rosenberg's reasoning and his conclusions on many different grounds. Perhaps his most telling criticism was that Rosenberg should have tried to reproduce the inverse incentive effects previously attributed to dissonance theory by adding conditions to his design in which the same experimenter called for the essay and measured the subsequent attitude. Aronson argued that there were many differences between the Cohen experiment and the Rosenberg replication, and to assume that his results reversed the dissonance proposition solely because evaluation apprehension was removed is unwarranted.

The fact remains that Rosenberg (1965) was able to obtain a positive relationship between amount of incentive and inferred attitude change, and the intriguing empirical and theoretical problem is how to account for the fact that both dissonance and reinforcement effects have been found within the forced-compliance paradigm. Carlsmith, Collins, and Helmreich (1966) have predicted and successfully produced these opposing effects in a context approximating the original Festinger and Carlsmith (1959) study. When the subject was induced to describe the task as attractive to the next "unsuspecting subject," the former's subsequent task-attractiveness ratings were more positive in low- than in high-incentive conditions. When the subject was instead asked to write an essay praising the task, portions of which might later be used by the experimenter, rated attractiveness varied directly with the amount of incentive offered. Carlsmith et al. argued that amount of incentive will relate directly to attitude change (a reinforcement effect) whenever the dissonance involved in a counterattitudinal act is minimal. The subject who complied in the essay-writing conditions of their experiment had a number of legitimate reasons for doing so, and the experimenter, the only person to read the essay, knew full well that the essay did not express the subject's private opinion. Dissonance should be much greater in the conditions requiring the subject to dupe another person like himself.

The Carlsmith et al. (1966) experiment is especially important because of the care with which it was conducted, the clear replication of the Festinger and Carlsmith (1959) results it provides, and the separate elicitation of both dissonance and reinforcement effects within the same general design. But while they may account for the Rosenberg reinforcement effect, Carlsmith et al. are left without any clear explanation for Cohen's (1962) results. After all, he required an essay rather than a deceitful confrontation with another subject and obtained dissonance rather than reinforcement effects. One could argue that attitudes toward the New Haven police are likely to be more central and important than attitudes toward a boring task, and thus a counterattitudinal essay is more inherently dissonant in the former case. Or, one could argue that the subjects in Cohen's experiment were not really assured anonymity (as in Carlsmith et al.'s). Nevertheless, the empirical discrepancies existing in the forced-compliance literature are not entirely reconciled by the Carlsmith et al. study.

The major focus of these studies has been the relationship between the amount of incentive offered and subsequent attitude change, but a clear prediction from dissonance theory cannot be made unless the subject makes his decision to comply *after* considering the incentive magnitude. The incentive must be one of the conditions potentially affecting the decision to comply rather than a reward for having already so decided.

Both Cohen (1962) and Rosenberg (1965) reported that they took care to assure subjects that the decision to write the essay was entirely their own. It may be argued, however, that Rosenberg's major alteration of Cohen's procedure, the separation of the compliant-behavior setting from the attitude-measurement setting to eliminate evaluation apprehension, reduced his subjects' freedom not to comply. When Rosenberg's subjects arrived for the experiment, they found him busily engaged and were given the option of waiting for "15 or 20 minutes" or, as an afterthought, participating in "another little experiment some graduate student in education is doing." Professing to know little about this other experiment except that it "has to do with attitudes" and "I gather they have some research funds," Rosenberg did not pressure the subject into a decision, but let him decide for himself whether he wanted to participate or wait. Having made the decision to participate, each subject further strengthened his commitment by walking to the location of the second experiment. The choice then offered by the second experimenter was considerably less than a free one. Being already effectively committed, the subject would be more likely to treat the subsequent monetary offering as a bonus for prior compliance than as one of the conditions to be considered in making a free choice.

If the preceding argument is correct, Rosenberg's findings cannot be compared with Cohen's because different conditions prevailed in the two experiments when the counterattitudinal essays were written. Rosenberg inadvertently made it difficult for subjects not to comply and found that degree of attitude change was positively related to incentive magnitude, in support of a reinforcement position or an affective-cognitive consistency model (Rosenberg, 1960). In contrast to this, Cohen's procedure presented the choice not to comply as a more viable alternative and found that attitude change was inversely related to incentive magnitude, in support of a derivation from dissonance theory. A meaningful resolution of these discrepant findings would be to show that the effects of incentive magnitude on attitude change are either direct or inverse, depending on the presence or absence of freedom not to comply. The first experiment to follow was conducted as a direct test of the role of such freedom to choose not to engage in counterattitudinal behavior.

EXPERIMENT I

Method

Attitude issue and subjects. At the time of the first experiment a rather heated controversy was raging in the state of North Carolina concerning the wisdom of a law that forbade Communists and Fifth Amendment pleaders from speaking at state-supported institutions. On the basis of informal opinion sampling, fortified by the plausible expectation that students deplore implied restrictions on their own freedom to listen, we assumed that college-student subjects would be strongly opposed to speaker-ban legislation. The issue thus seemed comparable to "the actions of the New Haven police" (Cohen, 1962) and to a ban on Ohio State's participation in the Rose Bowl (Rosenberg, 1965).

Fifty-five introductory psychology students at Duke University served as subjects in the experiment. Forty subjects (15 males and 25 females) were randomly assigned to four experimental conditions[2]; 13 were subsequently assigned to a control condition. All experimental subjects were asked to write a "forceful and convincing essay" in favor of the speaker-ban law. After writing the essay, each subject was asked to indicate his opinion about the speaker-ban law by checking a point on a 31-point scale comparable to Cohen's (1962) and Rosenberg's (1965) measure. The scale read, "In my opinion the Speaker Ban Law of North Carolina is . . . ," followed by 31 horizontal dots with seven labels ranging from "not justified at all" to "completely justified." Subjects in the control condition merely filled out the scale without having previously written a pro-speaker-ban essay.

Procedure and Design

The basic procedure was closely modeled after that of Cohen (1962) except that the subjects were recruited from the introductory psychology course and came individually to the laboratory, rather than being approached in their dormitory rooms. The experimenter introduced himself as a graduate student in psychology. In the *free-decision condition* he immediately said, "I want to explain to you what this task is all about. I want to make it clear, though, that the decision to perform the task will be entirely your own." In the *no-choice condition* he merely said, "I want to explain to you what this task that you have volunteered for is all about." He then proceeded in both conditions to provide the following rationale for the essay-writing task:

---

[2] Two more experimental subjects were actually run whose data were not analyzed. One of these was obviously in favor of the speaker-ban law at the outset, and the other was the victim of experimenter error in presenting instructions.

> The Association of Private Colleges of the Southeast, of which Duke is a member, is considering the adoption of a uniform speaker policy that would be binding on its member schools. Before they can decide what kind of policy to adopt, if indeed they decide to adopt one, they have undertaken a large scale research program in order to help them understand what the issues really are. This study is part of that program. The APCSE is working through the Department of Psychology here at Duke and through the departments of psychology at other private schools in the area because of the access which the department has to a wide cross-section of students such as yourself who must participate in psychological experiments and because of the number of graduate students that are available to conduct research. We have found, from past experience, that one of the best ways to get relevant arguments on both sides of the issue is to ask people to write essays favoring only one side. We think we know pretty much how you feel about the student's rights in this matter. [Here the experimenter paused and waited for a comment that would confirm the subject's initial opinion opposing the speaker-ban law. Only one subject expressed a favoring opinion at this point; see Footnote 2.] Nonetheless, what we need now are essays favoring the speaker ban.

At this point, the free-decision and no-choice conditions again diverged. In the free-decision condition the subjects were told that the APCSE was paying $.50 (low incentive) or $2.50 (high incentive) in addition to the standard experimental credit given to all subjects. The experimenter again stressed that the decision to write the essay was entirely up to the subject and that he would receive experimental credit in any case. In the no-choice condition the experimenter acted as if, naturally, the subject in volunteering for the experiment had committed himself to its requirements. He simply pointed out that the experiment involved writing a strong and forceful essay favoring the speaker-ban law. After the subject was handed a pencil and some paper, but before he began to write, the experimenter broke in: "Oh yes, I almost forgot to tell you. . . . The APCSE is paying all participants $.50 [or $2.50] for their time."

In all conditions subjects were paid, *before* they wrote the essay, the amount of money promised them. The experimenter then left the room and allowed the subject 20 minutes to complete his essay. When he returned, the experimenter collected the essay, administered the brief attitude scale, and interviewed the subject concerning his perceptions of the experiment. No subject indicated any suspicion regarding the true purpose of the experiment. The purpose was then explained to each in detail, and all deceptions were revealed. None of the subjects recalled having any doubts about the existence of the fictitious APCSE. Each subject was ultimately allowed to keep $1.50. Because they were made to realize that they were assigned by chance to the high-inducement condition, those who had initially received $2.50 were quite agreeable when asked to return $1 of their money. Subjects in the low-inducement condition were delighted to learn of their good fortune—that they would receive $1 more than they had bargained for.

Results

Before the results bearing on the central hypothesis are presented, it is of interest to note the difference in decision time in the two free-decision conditions. After the experimenter began to notice that *free-decision* subjects in the low-incentive condition took much longer to make up their minds about writing the essay than *free-decision* subjects in the high-incentive condition, he started to record decision times with a hidden stopwatch. The last seven subjects in the low-incentive condition took an average of 25.29 seconds to reach a decision; the comparable mean for the last seven subjects in the high-incentive condition was 11.00 seconds. In spite of the reduced n, this difference is significant ($p < .025$, U test). This evidence strongly suggests that there was greater predecisional conflict in the low-incentive condition, and thus the conditions are appropriate for testing the dissonance hypothesis: since predecisional conflict leads to postdecisional dissonance (Festinger, 1964), more dissonance and hence more attitude change should occur in the free-decision–low-incentive condition.

After establishing that the means for female and male subjects were nearly identical ($t = .18$), the posttreatment attitude scores were placed in a simple 2 (for Degree of Decision Freedom) × 2 (for Level of Incentive) factorial design. The means for each condition are presented in Table 1. Scale values could range from 1.0 (antispeaker ban) to 7.0 (prospeaker ban). Table 2 summarizes the analysis of variance and appro-

TABLE 1

ATTITUDE-SCALE MEANS OBTAINED IN THE FIVE CONDITIONS: EXPERIMENT I

	Incentive	
	$.50	$2.50
No choice	1.66[a]	2.34
Free decision	2.96	1.64
Control[b]	1.71	

Note.—$n = 10$ under both incentives for free-choice and free-decision conditions. For the control condition, $n = 13$.

[a] The higher the number, the more the speaker-ban law was considered to be justified.

[b] Since subjects in the control condition were all run after the main experiment was completed, the mean for this condition is presented only as an estimate of student opinion toward the issue in the absence of dissonance or incentive effects. The data from the control condition were not included in the statistical analysis.

priate orthogonal comparisons. The prediction that the amount of inferred attitude change would relate positively to inducement level in the no-choice conditions and negatively in the free-decision conditions is clearly confirmed ($F_{1, 36} = 8.70$; $p < .01$). The dissonance effect in the free-decision condition was itself significant; the reinforcement effect in the prior-commitment condition was not. The control subjects, who checked the scale without writing an essay, were about as much against the speaker ban as subjects in the conditions where little or no change was predicted.

In an effort to shed light on possible mechanisms underlying these findings, the essays themselves were examined. The average number of words per essay was 192.3, and there were no significant differences among the four conditions in essay length. The essays were evaluated in a manner similar to that described by Rosenberg (1965). Two independent raters, blind as to the subject's condition, rated the essays in terms of the degree of organization manifested and the degree of "intent to persuade." Each of these ratings was made in terms of a 5-point scale. The judges agreed or were 1 point discrepant in 72% of the organization ratings and 85% of the persuasiveness ratings. These percentages of agreement were comparable to those obtained by Rosenberg (1965), but two independent judges using 5-point rating scales should, by chance, be no more than 1 point discrepant on more than 50% of their ratings. When a more traditional estimate of the reliability of the ratings was calculated (Winer, 1962, pp. 124 ff.), it was found that the reliability coefficient for the ratings of degree of organization was .54, and the coefficient for the ratings of persuasiveness was .55. These coefficients estimate the reliability of the ratings that result from averaging over the two judges. When these ratings were submitted to an analysis of variance, there were no differences among conditions in either organization or persuasiveness.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE: EXPERIMENT I

Source of variation	*MS*	*F*
Choice (A)	0.90	<1
Incentive (B)	1.02	<1
A × B	10.00	8.70****
Error	1.15	
Low incentive vs. high incentive within free-decision conditions	8.71	7.57****
Low incentive vs. high incentive within no-choice conditions	2.31	2.01*

Note.—Two-way analysis of experimental conditions.
* $p < .20$, $df = 1/36$.
**** $p < .01$, $df = 1/36$.

Since the reliability of the ratings discussed above was quite low, an attempt was made to obtain ratings of acceptably high reliability. Two varsity debate partners agreed to rate the essays. General criteria to be used in determining the ratings were discussed, but the ratings were made independently. Each essay was rated for the persuasiveness of the presentation on a 7-point scale. Sixty percent of these ratings were no more than 1 point discrepant; the reliability coefficient was .48. (The chance percentage for agreements or 1-point discrepancies is 39% when a 7-point scale is used by two independent judges.) There were again no differences among the conditions in the rating received. Also, no between-condition differences appeared on the ratings made by any individual judge.

Discussion

The major purpose of the present experiment was achieved: to show that dissonance and reinforcement effects can be obtained within the same forced-compliance paradigm by varying the degree to which the subject is committed to comply before learning about the monetary incentive. Subjects who commit themselves after weighing the unpleasantness of the essay-writing task against the amount of incentive offered show the effects predicted by dissonance theory. The decision-time data strongly suggest that the subjects do in fact consider the essay-writing task unpleasant. Subjects who are not free to decide against compliance and then learn about a financial "bonus" produce results in line with reinforcement theory (that which is associated with something of value itself takes on value) or in line with the more complex affective-cognitive consistency model espoused by Rosenberg.

The present study was stimulated by Rosenberg's (1965) experiment, but the relevance of the results to a critique of Rosenberg's conclusions rests on the claim that his way of removing evaluation apprehension precommitted the subject to an unpleasant task before he had a chance to weigh the incentive for compliance. If this criticism is valid, then it should be possible to reproduce Rosenberg's results by closely replicating his procedures, and to obtain the converse of these results (confirming the dissonance prediction) by insuring that the subject does not commit himself before being confronted with the incentive for compliance. A second experiment was planned in an attempt to do precisely this.

EXPERIMENT II

Method

Attitude issue. As we prepared to run the second experiment, certain paternalistic policies of the Duke University administration were being challenged by the undergraduates, and there was a movement toward liberalization of *in loco parentis* social regulations. It was assumed, therefore, that undergraduates who were induced to write forceful and convincing essays in support of strict enforcement of *in loco parentis* policies would be performing a counterattitudinal task.

Subjects. Fifty-nine male introductory psychology students volunteered to participate for experimental credit in a study described as an "Attitude Survey." The data of 50 of these students, who were randomly assigned to the four experimental conditions and the control condition, were used in the reported analysis. Six subjects were eliminated because they did not complete the experimental procedure. Usually, they chose to read or study while waiting for the first experimenter rather than go to the second experimenter. Another subject was eliminated because he was initially in favor of strict *in loco parentis* policies, and writing the essay would not have been counterattitudinal for him.

Only two subjects who had completed the procedure were eliminated from the analysis. The first of these was excluded when it was discovered during the final interview that he had misinterpreted the attitude questionnaire. The second was eliminated because he accurately perceived the true purpose of the experiment. Both subjects had been assigned to the *free-decision–high-incentive condition.* The results of the study are not changed if these two subjects are included in the analysis.

Procedure and design. The procedure was a close approximation to that used by Rosenberg (1965). All subjects reported to the office of the first experimenter (E_1) where they found E_1 engaged in conversation with another student and were told, "I'm sorry, but I'm running late on my schedule today, and I'll have to keep you waiting for about 15 or 20 minutes. Is that all right?" All subjects agreed to wait.

Each experimental subject was then told:

> Oh, I've just thought of something; while you're waiting you could participate in another little experiment that some graduate student in education is doing. This fellow called me the other day and said he needed volunteers in a hurry for some sort of study he's doing—I don't know what it's about exactly except that it has to do with attitudes and that's why he called me, because my research is in a similar area as you'll see later. Of course, he can't give you any credit but I gather thay have some research funds and they are paying people instead. So, if you care to do that, you can.

At this point, one-half of the experimental subjects (*prior-commitment condition*) were allowed to leave for the second experiment without further comment by E_1. Since it was believed that Rosenberg's procedure restricted subjects' freedom not to comply with the task of the "little experiment," it was decided to manipulate degree of choice by removing this restriction. Thus, for subjects in the *free-decision* condition, after the subject had agreed to participate in the second experiment, E_1 added:

> All I told this fellow was that I would send him some subjects if it was convenient but that I couldn't obligate my subjects in any way. So, when you get up there, listen to what he has to say and feel free to decide from there.

All experimental subjects then reported to the second experimenter (E_2). To control for the effects of experimenter bias, E_2 was not informed whether the subject was in the prior-commitment condition or the free-decision condition. E_2 presented himself as a graduate student in the Department of Education and introduced the essay-writing task using a procedure that, as in Experiment I, very closely approximated Cohen's (1962).

Rather than the free-decision versus no-choice manipulation of Experiment I, E_2 began by saying to all subjects, "At the present time, Duke University is beginning to question the wisdom of assuming the role of 'substitute parent' to its students." From that point, the instructions paralleled those of Experiment I with the substitution of *in loco parentis* regulations for the speaker-ban law. After confirming that the subject held an opinion opposed to rigid *in loco parentis* regulations, E_2 concluded:

> What we need now are essays favoring a strict enforcement of *in loco parentis*. So, what we would like you to do—if you are willing[3]—is to write the strongest, most forceful and most convincing essay that you can in favor of a strict enforcement of the substitute parent concept here at Duke.

It was then explained that the sponsoring agency was offering either $.50 (*low-incentive* conditions) or $2.50 (*high-incentive* conditions) for participation in the study. When the subject agreed to write the essay, he was paid the money promised to him and then began the task.

After completing the essay and being thanked and dismissed by E_2, all experimental subjects returned to E_1's office. To introduce the dependent measure, E_1 explained:

> What I had wanted you to do was participate in a continuing study I carry on every semester as a sort of Gallup poll to keep a check on opinion patterns on different University issues. I'd like you to fill out this questionnaire as an objective indication of your opinions and when you've finished I'd like to chat for a while about various issues on campus. OK?

E_1 was not informed of the amount of money the subject had received, and in no case did he find out until after the subject had completed the dependent measure.

The dependent measure consisted of an eight-item questionnaire dealing with various university issues. The critical item read, "How justified is the Unversity's policy of assuming parental responsibilities for its students?" and was accompanied by the familiar 31-point scale. When the subject had completed the questionnaire, E_1 put it aside (without looking at the responses) and began a structured interview that included probes for suspicion and checks on perceptions of the manipulations. When E_1 was satisfied that the subject had not perceived the true purpose of the experiment, he revealed the deceptions and explained the necessity for them. As in Experiment I, all experimental subjects agreed to accept $1.50 for their time.

Subjects assigned to the control condition also found E_1 engaged in conversation and were asked if they could return in 15 or 20 minutes. Upon their return they were treated exactly as experimental subjects.

These procedures resulted in five conditions: two levels of incentive magnitude under a condition of free decision, the same two levels under a condition of prior commitment, and the control condition.

Results

The mean attitude-scale scores on the critical item for each of the five conditions are presented in Table 3. It can be seen that the results are very similar to those of Experiment I. The data were submitted to a one-way analysis of variance, summarized in Table 4. The overall treatment effect was significant ($F_{4,\,45} = 4.02$; $p < .01$). The two comparisons reflecting the hypotheses of this study were also significant: Within the free-decision conditions a low incentive produced more inferred attitude change than a high incentive ($F_{1,\,45} = 6.82$; $p < .025$). Within the prior-commitment conditions this effect was reversed, and a high incentive produced more inferred attitude change than a low incentive ($F_{1,\,45} = 4.90$; $p < .05$). The position of the control group indicates that differences between the experimental conditions resulted from positive attitude change rather than a combination of positive and negative changes.

Once again we attempted to investigate the possibility that these effects were mediated by some aspect of the counterattitudinal performance. Two raters, working independently and without knowledge of the experi-

TABLE 3

ATTITUDE-SCALE MEANS OBTAINED IN THE FIVE CONDITIONS: EXPERIMENT II

	Incentive	
	$.50	$2.50
Prior commitment	2.68[a]	3.46
Free decision	3.64	2.72
Control	2.56	

Note.—$n = 10$ in all conditions.
[a] The higher the number, the more strict application of *in loco parentis* regulations was considered justified.

---

[3] This vague statement of choice was given to all subjects in order to keep the instructions constant across experimental groups and to enable E_2 to remain "blind" as to the condition of each subject. It was assumed that the crucial manipulation of free decision versus prior commitment had already been accomplished by E_1.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE: EXPERIMENT II

Source of variation[a]	*MS*	*F*
Treatment	2.49	4.02****
Error	.62	
Low incentive vs. high incentive free-decision conditions	4.23	6.82***
Low incentive vs. high incentive prior-commitment conditions	3.04	4.90**

Note.—One-way analysis of five conditions.

[a] The control condition differs significantly from both the prior-commitment–high-incentive and the free-decision–low-incentive conditions.

** $p < .05$, $df = 1/45$.

*** $p < .025$, $df = 1/45$.

**** $p < .01$, $df = 4/45$.

mental conditions, rated each essay on 7-point scales for the extremity of attitudinal position advocated, the persuasiveness of the essay, and its degree of organization. The two raters agreed or were within 1 point of each other for 65% of the essays when estimating the attitudinal position, 60% when rating them for persuasiveness, and 52.5% when rating them for organization. The reliability coefficient for the estimated attitudinal position (Winer, 1962, pp. 124 ff.) was .67, the coefficient for the persuasiveness ratings was .51, and the coefficient for the organization ratings was .38. There were no differences among conditions on any of these ratings. The essays were then rated for the persuasiveness of the presentation on a 7-point scale by the same varsity debaters as had rated the essays from Experiment I. The debaters agreed or were within 1 point of each other for 65% of the essays, and the reliability coefficient was a somewhat more acceptable .71. Again, however, there were no differences among conditions on these ratings, whether the judges' ratings were averaged or each judge's ratings were examined separately. In a final attempt to find a performance difference among the conditions the number of words in each essay was counted; the conditions were compared on this measure of performance and were found not to differ from one another.

Discussion

The results of Experiment II support the argument that Rosenberg's (1965) procedure for the elimination of evaluation apprehension committed his subjects to perform the essay-writing task before they learned of the nature of the task and the amount of reward offered. The positive relationship between incentive magnitude and attitude change in the prior-commitment conditions of the present experiment replicates the no-choice results of Experiment I and the relationship found by Rosenberg (1965). It could be argued on this basis alone that such procedures as Rosenberg's have the same effect as allowing the subject no choice concerning performance of the counterattitudinal act. The argument becomes much more convincing, however, if it can be shown that appropriate alteration of Rosenberg's procedure, reducing the prior commitment of the subject, leads to an *inverse* relationship between incentive magnitude and attitude change. The free-decision conditions of Experiment II demonstrate precisely this point: when the subject does not feel that he has previously committed himself to performance of the counterattitudinal action requested by E_2, attitude change is an inverse function of incentive magnitude.

It should be noted here that a "balanced replication" (Aronson, 1966) of Rosenberg's (1965) study was required. Had Experiment II included only the free-decision conditions it would be possible to argue that our procedure was not successful in eliminating evaluation apprehension and that the results reflected once again the effect of this contaminant in research on forced compliance. The results of the prior-commitment conditions of Experiment II, however, counter this criticism. A persistent critic might still argue that the free-decision manipulation reintroduced evaluation apprehension. Perhaps the comment added to create the free-decision condition in some way increased the chances that subjects would see the experiments as related. However, the structured interview conducted by E_1 revealed no differential level of suspicion between the prior-commitment and free-decision conditions. In the absence of a reliable and valid measure of evaluation apprehension, we can do no more than contend that our interview was sensitive enough to detect suspicion and that we found no indication of differential suspicion among the conditions.

The results of the two studies reported above imply that the discrepancy between Cohen's (1962) findings and the results of Rosenberg's (1965) experiment may indeed be resolved in the manner indicated earlier in this paper. For Cohen's subjects the decision not to comply was a viable alternative at the time they were confronted with the essay-writing task and offered an incentive of certain value. Under such conditions dissonance will be induced whenever the incentive is not large enough to justify performance of the task, and incentive magnitude will be inversely related to subsequent attitude change. However, if a subject's freedom not to comply has been restricted before he is confronted with the task and with a clear description of the incentive, dissonance cannot be induced by an incentive of insufficient magnitude. Under these conditions, the reinforcing properties of the incentives will lead to a positive relationship between incentive magnitude and attitude change. Although Rosenberg (1965) demonstrated such a relationship, his assertion that it may be obscured by failure to remove evaluation apprenhension seems no longer tenable. No attempt was made in the procedure of Experiment I to remove evaluation apprehension, and yet the results are very similar to the results of Experiment II.

Rosenberg (1966) has more recently advanced two additional hypotheses intended to resolve discrepancies in the forced-compliance literature. The first of these is that we must distinguish counterattitudinal actions that are simple and overt from those featuring the elaboration of a set of arguments. Supposedly a performance of the former kind (e.g., eating a disliked food) will lead to the inverse relationship between attitude change and incentive magnitude, while an act of the latter kind (e.g., writing a counterattitudinal essay) will result in a positive relationship. The second hypothesis is that we must distinguish between two kinds of counterattitudinal performances: (*a*) those carried out under instructions that lead the subject to believe his performance will be used to deceive others, and (*b*) those following from instructions to elaborate, for some reasonable and legitimate purpose, a set of arguments opposite to his private opinion. It is hypothesized that even if the actual task is the same, say essay writing, the first type of instruction will lead to an inverse relationship between incentive magnitude and attitude change, and the second type of instruction will lead to a positive relationship.

In the studies reported above the subject's task was presented with no hint that his performance would be used to deceive anyone, and the task in all cases was to elaborate a set of arguments opposite to his own opinion. It follows from the two hypotheses suggested by Rosenberg (1966) and presented above that we should not have been able to obtain the inverse relationship between incentive magnitude and attitude change using our procedures. However, in both experiments, we obtained the positive *and* the inverse relationship. We are forced to conclude that neither the "simple versus complex" hypothesis nor the "duplicity versus legitimate" hypothesis can account for the present results.

In place of these hypotheses we conclude that at least some of the discrepancies in the forced-compliance literature may be resolved by closer attention to the role of decision freedom at the time the incentive is offered. A barely sufficient incentive for making counterattitudinal statements *does* result in dissonance and subsequent attitude change if the subject feels he is quite free not to comply. When the freedom not to comply is removed or markedly decreased, on the other hand, attitude change is greater the greater the incentive for compliance.

REFERENCES

ARONSON, E. The psychology of insufficient justification. In S. Feldman (Ed.), *Cognitive consistency: Motivational antecedents and behavioral consequents*. New York: Academic Press, 1966. Pp. 115–133.

CARLSMITH, J. M., COLLINS, B. E., & HELMREICH, R. L. Studies in forced compliance: I. The effect of pressure for compliance on attitude change produced by face-to-face role playing and anonymous essay writing. *Journal of Personality and Social Psychology*, 1966, 4, 1–13.

COHEN, A. R. An experiment on small rewards for discrepant compliance and attitude change. In

J. W. Brehm & A. R. Cohen, *Explorations in cognitive dissonance*. New York: Wiley, 1962. Pp. 73–78.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L. *Conflict, decision, and dissonance*. Stanford: Stanford University Press, 1964.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology,* 1959, **58**, 203–210.

NUTTIN, J. M., JR. Dissonant evidence about dissonance theory. Paper read at 2nd conference of experimental social psychologists in Europe, Frascati, Italy, 1964. (Mimeo)

ROSENBERG, M. J. An analysis of affective-cognitive consistency. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change*. New Haven: Yale University Press, 1960. Pp. 15–64.

ROSENBERG, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology,* 1965, **1**, 28–42.

ROSENBERG, M. J. Some limits of dissonance: Toward a differentiated view of counter-attitudinal performance. In S. Feldman (Ed.), *Cognitive consistency: Motivational antecedents and behavioral consequents*. New York: Academic Press, 1966. Pp. 135–170.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received August 1, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 255–263

BEHAVIORAL CORRELATES OF ACADEMIC ACHIEVEMENT:

CONFORMITY UNDER ACHIEVEMENT- AND AFFILIATION-INCENTIVE CONDITIONS[1]

ROBERT S. WYER, Jr.
University of Illinois, Chicago Circle

61 freshman males and 64 freshman females, representing 4 combinations of academic aptitude (college entrance examination score) and performance (1st-term grade-point average), were asked to estimate the number of dots on slides before and after exposure to fictitious group judgments. Conformity was measured under achievement-incentive conditions, in which Ss were told that their performance on the task would reflect their achievement potential, and affiliation-incentive conditions, in which the importance of the task was deemphasized and group attractiveness was increased. Low performers of low ability were relatively high in conformity, while low performers of high ability (underachievers) were low. These relationships were more pronounced when an incentive to perform well was provided. When an affiliation incentive was provided, conformity was related negatively to performance among females but increased with performance among males. Similar relationships occurred between academic aptitude and conformity. High-performing males conformed less under achievement conditions than under affiliation conditions, while high-performing females conformed less under affiliation conditions than under achievement conditions. High-performing males conformed more than high-performing females under affiliation conditions, but somewhat less than female high performers when an incentive to perform well was provided. Among low performers, similar relationships were not detected. Results provided insight into the motivational correlates of academic performance and aptitude among male and female college students, and into the behavior typical of these students in achievement and social situations.

Although studies have generally found support for the hypothesis that social dependence and conformity are related negatively to academic achievement among college students, this research has been difficult to interpret for at least three reasons. First, academic performance has frequently been measured without controlling for academic aptitude (cf. Gilmore, 1951). Second, conformity has typically been inferred either from subjects' acknowledgment of this behavior on questionnaires (Merrill & Murphy, 1959; Weigand, 1953) or from projective techniques (Burgess, 1956); the degree to which such measures reflect behavior in actual social situations is unknown. Third, little attention has been given to situational differences in the motives for conformity.

The research reported here, one of a series of studies on nonintellective factors associated with academic achievement among college students (Wyer, 1965; Wyer & Terrell, 1965; Wyer, Weatherley, & Terrell, 1965), was designed to help clarify these issues. Representatives of four combinations of academic aptitude (inferred from college entrance examination scores) and academic performance (grade-point average) made judgments of physical stimuli before and after exposure to fictitious group norms. Conformity to group estimates was analyzed as a function of aptitude, performance, and sex under each of two incentive conditions. In one condition, performance on the task was associated with achievement potential, and subjects were told that their scores would be made part of their academic records. In the second condition, the importance of the task was deemphasized; however, an attempt was made to increase the value attached to favorable social relationships among group members. This manipulation of achievement incentive versus affiliation incentive was particularly relevant in light of earlier research indicating possibly different effects of achievement and social motives upon performance among males and females (Wyer & Terrell, 1965) and in view of the different motives underlying conformity (cf. Asch, 1951; Becker, Lerner, & Carroll, 1964; Deutsch & Gerard, 1955; Kelman, 1961; Wyer, 1966).

---

[1] This research was sponsored by the United States Office of Education, Cooperative Research Project No. 5-1008, under the direction of Glenn Terrell. Thanks are extended to Thomas Gronek, Douglas Johnson, and Patricia Creedon for their assistance in various phases of this research, and to Harry Upshaw for his helpful criticisms of the manuscript.

A student's academic aptitude, as the term is used in this study, is assumed to indicate his mastery of general intellectual skills before entering college, and therefore to reflect the level of performance he should attain with an average amount of effort. Overachievers and underachievers are considered to be students whose performance exceeds or falls short of the level they would be predicted to attain with an average amount of effort.

Two conditions may be necessary for conformity to occur on a task that measures personal achievement: First, the appearance of competence in achievement-related activity must be important to the subject; second, the subject must have reason to believe that the judgments of other group members are more likely to be correct than his own. If these assumptions are valid, the degree to which students at different levels of aptitude and performance conform under these conditions may allow inferences to be made concerning both their desire to perform well in achievement situations and their tendency to rely upon others for assistance in this activity.

Because of their previous history of failure in achievement situations, students who both are low in measured aptitude and have performed poorly in college have the greatest reason to believe that their peers' performance in such situations is likely to be superior to their own. These students were therefore expected to conform more under achievement-incentive conditions than students who, as a result of either good grades or high measured ability, have received some indication that they are competent.

In contrast to low performers of low ability, high-aptitude students who perform poorly in college (underachievers) may be quite confident of their ability to perform well. Their low grades may therefore indicate that they devote less effort to the pursuit of academic goals than is typical of college students. This may be because they attach little importance to demonstrating their competence in achievement situations and even may actively avoid doing so. Or, they may be generally independent and rebellious. In any case, when the primary reason for conformity is to appear competent in achievement activity, the conformity of these students should be low.

Students who perform well in college despite low measured ability (overachievers) presumably devote greater effort than the typical college student to the pursuit of academic goals, probably as the result of a high desire to achieve. However, whether or not these students tend to rely upon others for assistance in attaining achievement goals, and therefore whether they would conform to peers' judgments in achievement situations, was unclear.

When the importance of personal achievement was deemphasized and group attractiveness was increased, the primary motive for conformity was assumed to be an unwillingness to appear deviant in opinions from other group members, regardless of the validity of these opinions. Again, two things may be necessary for conformity to be appreciable in such a situation: First, the subject must have some interest in maintaining favorable social relationships with other group members; second, he must have reason to believe that the quality of these relationships can be improved by conforming. In an experimental situation comparable to the one used in this study, Wyer (1966) found that conformity was appreciable only among subjects who were led to believe they were rejected by group members to whom they were attracted. Persons who were told they were accepted by persons they liked, and therefore were given reason to believe that their group status was secure, conformed no more than did subjects who disliked other group members. Similar relationships involving group acceptance have been reported by Harvey and Consalvi (1960) and Dittes and Kelley (1956).

These findings have implications for the relationships of academic variables to conformity under similar conditions. Wyer and Terrell (1965) concluded that academic-achievement-related activity is detrimental to social goal attainment among males. If this is true, high-performing males, who presumably manifest academic-achievement-related activity, should be more uncertain of their social acceptance than low-performing males, and therefore should be relatively more conforming when the importance of achievement is low and their attraction to the group is high. On the other hand, Wyer and Terrell assumed that academic achievement does not conflict with social goal attainment among female students, but instead may actually facilitate the attainment of these goals. To this extent, females who perform well should have relatively more confidence in their peer-group acceptance than females who perform poorly, and therefore should tend less to conform merely to avoid appearing deviant from other group members.

While the conclusions drawn in the Wyer and Terrell study allow fairly clear predictions to be made concerning the relationships between academic performance and conformity under affiliation-incentive conditions, the earlier study unfortunately failed to provide convincing support either for these conclusions or for the reasons underlying them. The results of the present study were expected to provide evidence concerning the merits of these conclusions. Furthermore, by considering behavior in both achievement- and affiliation-incentive conditions, it was hoped that the reasons underlying sex differences in the relationship between academic and social variables could be clarified.

METHOD

Subjects were selected from 2,000 second-quarter freshmen at a midwestern urban university for whom indexes of academic aptitude and performance were available. The subpopulation comprised the majority of the freshman class and was therefore representative of the freshman student body. Academic aptitude was measured by composite score on the American College Testing Service college entrance examination. Performance was measured by first-term grade-point average. In the population from which subjects were drawn, sex differences in aptitude (males, $M = 24.0$; females, $M = 23.3$) and in performance (males, $M = 3.06$; females, $M = 3.20$) were not appreciable. Measures of aptitude and performance were each converted to z scores, and subjects were placed into four combinations of aptitude and performance according to the following criteria:

1. High aptitude, high performance—z scores of greater than .50 for both aptitude and performance, and an absolute difference between aptitude and performance z scores of less than .30.
2. Low aptitude, low performance—z scores of less than $-.50$ for both aptitude and performance, and an absolute difference between aptitude and performance z scores of less than .30.
3. High aptitude, low performance—an aptitude z score of greater than .50, a performance z score of less than $-.50$, and a difference between aptitude and performance z scores of greater than 1.30.
4. Low aptitude, high performance—an aptitude z score of less than $-.50$, a performance z score of greater than .50, and a difference between performance and aptitude z scores of greater than 1.30.

Sixteen males and 16 females were selected at each of the four combinations of aptitude and performance and recruited for the experiment. They were each paid $1 for their services.

Sixteen groups of from 7 to 10 subjects each were constructed. Each group was homogeneous with respect to sex. Groups were arranged so that eight males and eight females representing each of the four combinations of aptitude and performance would make estimates of the number of dots presented on slides under each of the two incentive conditions described below. Conformity under each condition was defined as the percentage of increase in estimates after exposure to fictitious group norms.

Manipulation of Incentive

Achievement incentive. To provide an incentive to perform well, accuracy on the judgmental task was associated with competence in achievement-related activity. Subjects were introduced to the task with the following instructions:

> This study is part of some important research sponsored by the federal government. We are investigating factors that affect achievement both in college and in later life. Knowledge of these factors when persons are in college may help us to predict how successful they will be later on. One characteristic we think may contribute to success is the ability to make judgments rapidly and accurately. Previous research indicates that persons who do well on the task you are about to perform are often able to make quick and accurate decisions under pressure, while those who perform poorly may not be as able to be counted upon when quick decisions are required. The Office of Admissions and Records has in fact asked that your scores on this task be made part of your permanent record.

Affiliation incentive. To construct a situation conducive to normative influence, the importance of the task was minimized. Furthermore, since group attraction seemed to be a necessary condition for this influence to be effective (Wyer, 1966), an attempt was made to increase this attraction. As part of an earlier study, all subjects had been administered a questionnaire dealing with attitudes and values. To increase group attraction, reference was made to this questionnaire:

> The study you are about to participate in is part of some research sponsored by the government. By the way, you may recall that at the deans' meetings this past fall each of you was asked to fill out a questionnaire dealing with general attitudes and values. In selecting students to take part in this subsequent study, we attempted to group persons whose values were similar to one another and therefore should be able to get along well together. In this particular group we were especially fortunate. The values of each of you were found to be very similar to those of others in the group. This means that, although you may not know one another at the present time, if you did you would probably find that you would get along very well together.
>
> The first thing I would like you to do this afternoon is to estimate the number of black dots presented on each of a set of slides. Since this is the first time we have used this task, we are really interested only in determining whether it will be useful in other experiments we are performing.

Assessment of Conformity

A procedure similar to that described by Wyer (1966) was used. Subjects in each experimental condition were asked to estimate the number of randomly arranged dots presented on each of 10 slides projected sequentially on a screen for 1.5 seconds each. About 5 seconds between presentations of successive slides were allowed for subjects to record estimates. A trial consisted of one presentation of the 10 slides. The order of presentation was the same for each of two trials prior to exposure to fictitious group norms and one trial following exposure to these norms. The number of dots on each slide, in order of their presentation, was 20, 37, 28, 40, 32, 20, 37, 28, 40, and 32. Estimates on Trials 1 and 2 were recorded on sheets consisting of two columns of blanks, one column for each trial.

After instructions designed to induce differences in incentive, subjects were told the nature of the slides to be presented and that these slides would be shown for only a short period of time. They were told that after each slide was presented they should record their estimates on their answer sheets.

When all 10 slides had been presented, subjects were asked to make new estimates, as "accuracy may improve with practice." They were told that before they were shown each slide they should note the number of dots they had estimated during Trial 1 and then, when the slide was presented, determine whether their first estimate was correct or if they should change their estimate. Two trials were used to minimize the degree to which change after exposure to group estimates could be attributable to factors other than this influence.

After each judgment on Trial 2, subjects were also asked to estimate their certainty that their estimate was accurate by placing a check along a 5-point scale from 1 (extremely certain) to 5 (not at all certain).

Following Trial 2, answer sheets were collected and subjects given an interpolated task. During this period new answer sheets were prepared. For each subject, the estimates he had made during Trial 2 were copied by the experimenter in the first column of the answer sheet. In the second column, the experimenter recorded fictitious "group averages" which were prepared as follows: For each of slides 1, 4, 7, 8, and 10, the "average" deviated from the subject's second estimate for that slide by one dot. For each of slides 2, 3, 5, 6, and 9, the "average" exceeded the subject's second estimate for that slide by 30%.

Subjects were then provided with appropriate answer sheets and given the following instructions:

> Now I would like you to make some more estimates of the number of dots on the slides I showed you before. Although I cannot tell you exactly how well you performed the first time, I can give you some idea of how close each of your guesses was to what other persons in the group guessed. In the first column of your answer sheet I have written your last estimate of the number of dots on each slide. Beside each of these, in the second column, I have written about the number of dots which most of the other persons in the group guessed were on the slide. Although neither your guess nor others' guesses may be the actual number of dots on the slides, in this way you can see how close you were to what others guessed.

In addition to these instructions, subjects run under affiliation-incentive conditions were told that, after they made their judgments on this trial, their estimates would be written on the board so they could be compared with those made by other members. This additional instruction was expected to increase the magnitude of difference in susceptibility to normative influence.

Subjects were then told that before each slide was shown they should note both the number of dots they had guessed during Trial 2 and the number that others in the group guessed; then when the slide was shown, they should determine whether their last estimate was correct or if they should change their estimate. They were also asked to record the certainty that each judgment was accurate according to the procedure used during Trial 2.

After the experiment, subjects run under achievement-incentive conditions were informed that the information concerning the importance of the task was false. No attempt was made to debrief subjects run under affiliation conditions.

Scoring. Conformity was defined as the percentage of increase in estimates for slides 2, 3, 5, 6, and 9 from Trial 2 to Trial 3. It was determined for each subject by summing the differences between his estimates for these slides during Trial 3 and his estimates during Trial 2 and then dividing by the total number of dots he estimated were on these slides during Trial 2.

RESULTS

Before considering the relationship of academic variables to conformity it seemed advisable to determine the degree of difficulty of the judgmental task and its possible effects on conformity. Two factors were considered: subjects' estimates of the certainty of their judgments on the trial preceding exposure to group norms (Trial 2), and the actual accuracy of their judgments during this trial. Analyses of data pertaining to certainty yielded no significant effects. However, the mean rating of certainty following Trial 2 was 2.75 on the 5-point scale ranging from extremely certain to not at all certain. This was sufficiently near the center of the scale to justify the assumption that the task was neither so difficult that subjects would believe that they were merely guessing, nor so easy that they would be unlikely to try to improve their performance under achievement-incentive conditions.

The accuracy of each subject's estimates during Trial 2 was computed by summing the absolute differences between his response to each slide and the actual number of dots shown on the slide. Analyses of variance performed on these data yielded no significant differences under achievement-incentive conditions. Under affiliation-incentive conditions, accuracy was greater among high-performing males (mean discrepancy per slide $= 6.7$) than among low performers or among females ($M_{\text{res}} = 9.6$; $p < .05$). However, the correlation of these discrepancies and conformity ($r = -.13$) was sufficiently small to eliminate concern over the effect of initial accuracy on the degree of subsequent conformity; in any event, the direction of the correlation indicated that the lower the discrepancy (i.e., the higher the accuracy), the greater the subsequent conformity.

Conformity under the two incentive conditions as a function of aptitude, performance, and sex is shown in Table 1. Three subjects did not appear for the experiment: one low-aptitude, high-performance male under achievement-incentive conditions; and one low-aptitude, low performance male and one low-aptitude, high-performance male under affiliation-incentive conditions. All other cells contained eight subjects each.

A four-factorial analysis of variance of these data, performed using an unweighted means approximation procedure for unequal cell frequencies (Winer, 1962), is sum-

TABLE 1

CONFORMITY UNDER ACHIEVEMENT-INCENTIVE CONDITIONS AS A FUNCTION OF APTITUDE, PERFORMANCE, SEX, AND INCENTIVE CONDITIONS

	Males			Females			Combined sexes		
	Low performance	High performance	*M*	Low performance	High performance	*M*	Low performance	High performance	*M*
Achievement incentive									
High aptitude	4.97	8.70	6.83	5.27	9.30	7.28	5.12	9.00	7.06
Low aptitude	10.38	5.41	7.89	11.14	10.97	11.06	10.76	8.19	9.48
M	7.67	7.05	7.36	8.21	10.14	9.17	7.94	8.59	8.27
Affiliation incentive									
High aptitude	10.68	16.00	13.34	9.35	4.83	7.09	10.01	10.42	10.22
Low aptitude	8.84	8.34	8.59	11.92	7.58	9.75	10.38	7.96	9.17
M	9.76	12.17	10.97	10.64	6.20	8.42	10.20	9.19	9.70
Combined incentives									
High aptitude	7.82	12.35	10.08	7.31	7.06	7.19	7.56	9.71	8.64
Low aptitude	9.61	6.88	8.25	11.53	9.28	10.40	10.57	8.07	9.32
M	8.21	9.66	9.17	9.42	8.17	8.80	9.07	8.89	8.98

TABLE 2

Analysis of Variance Summary of Data in Table 1

Source	*MS*	*F*
Total	4.62	
Aptitude (A)	1.89	.47
Performance (B)	.13	.03
Sex (C)	.55	.14
Incentive (D)	7.95	1.99
A × B	21.45	5.38*
A × C	25.62	6.42*
A × D	11.96	3.00
B × C	4.58	1.15
B × D	2.75	.69
C × D	18.62	4.67*
A × B × C	6.92	1.73
A × B × D	3.35	.84
A × C × D	5.58	1.40
B × C × D	22.54	5.65*
A × B × C × D	.02	.00
Error	3.99	

* $p < .05$, $df = 1/109$.

marized in Table 2. Four interactions reached significance. It may be worthwhile to consider the nature of these effects before dealing with the specific hypotheses under investigation.

Data pertaining to the triple interaction of sex, incentive, and performance are of particular interest. Males conformed significantly more under affiliation-incentive conditions ($M = 11.0$) than under achievement conditions ($M = 7.4$; $F = 6.51$; $p < .05$). This relationship was particularly strong among males with high academic performance. Female high performers, on the other hand, conformed more under achievement conditions ($M = 10.1$) than under affiliation conditions ($M = 6.2$; $F = 3.81$; $p < .10$). Between-sex comparisons under each incentive condition indicated that when an affiliation incentive was provided, high-performing males conformed substantially more ($M = 12.2$) than did high-performing females ($M = 6.2$; $F = 8.91$; $p < .01$), while under achievement conditions they conformed somewhat less ($M = 7.1$) than did high-performing females ($M = 10.1$; $F = 2.39$; $p > .10$). No differences between male and female low performers were significant under either incentive condition.

The significant interaction of sex and academic aptitude was unexpected. This significance may be attributed primarily to the fact that conformity was related negatively to aptitude among females ($F = 5.18$; $p < .05$), but was related slightly positively to aptitude among males ($F = 1.71$; *ns*).

Investigation of the Aptitude × Performance interaction indicates that low-aptitude, low-performing students exhibited the greatest conformity ($M = 10.6$), while high aptitude, low performers conformed least ($M = 7.6$). The difference between these groups was significant ($F = 4.51$; $p < .05$). However, since these results were hypothesized to occur only when the task was associated with achievement potential, the failure for this interaction to be contingent upon incentive conditions was surprising. It should be noted that under achievement conditions conformity was less among high-aptitude, low-performance subjects than among representatives of each other combination of aptitude and performance, while low-aptitude, low-performance subjects conformed more than each other group. These relationships were apparent among both males and females. Under affiliation conditions, however, this was not the case. Furthermore, while a comparison is not statistically justified, the conformity of underachievers under achievement conditions ($M = 5.1$) was substantially less than that of underachievers under affiliation conditions ($M = 10.0$; $F = 5.98$; $p < .05$).

When an affiliation incentive was provided, conformity was hypothesized to increase with academic performance among males but to decrease with performance among females. Simple effects analyses of the significant triple interaction of sex, incentive, and performance support this hypothesis among females ($F = 4.84$; $p < .05$); the relationship was in the predicted direction among males but was not significant ($F = 1.44$). These relationships were not apparent when an incentive to perform well was provided. Under affiliation conditions, conformity was particularly pronounced among high-performing males of high aptitude ($M = 16.0$), but was particularly low among high-performing females of high aptitude ($M = 4.8$). These groups were significantly different from one another ($F = 15.72$; $p < .001$) and also from the residual of same-sex students run under these conditions (in each case, $F > 4.30$; $p < .05$).

DISCUSSION

While results tend to support the hypotheses made in this study, the relationships of conformity to the four independent variables considered were more complex than had been anticipated. The validity of the general conclusion, based upon questionnaire measures of dependence, that students who perform well academically are more socially independent than lower performers is called into serious question. While high performers of high ability appeared to be somewhat less conforming than low performers of low ability when their personal achievement was ostensibly being evaluated, they nevertheless were more conforming under these conditions than were low performers of high ability. Furthermore, under conditions in which personal achievement was not a factor, the degree of conformity of high performers was highly contingent upon sex.

In interpreting these and other results it is important to note that high-performing males conformed less under achievement-incentive conditions than under affiliation-incentive conditions, but high-performing females conformed more when personal achievement was emphasized than when the desirability of affiliative goals was increased. Since males as a group were expected to be relatively achievement oriented, and since this should be particularly true of males who perform well in college, it seems reasonable to attribute the low conformity of high-performing males under achievement-incentive conditions to a tendency not to depend upon others for assistance in achievement situations. When their personal competence in achievement activity is not involved, however, these students appear to be highly susceptible to social influence. If conformity in the latter situation increases when students believe they may not be accepted by other group members (Wyer, 1966), this may indicate that high-performing males are relatively uncertain of the quality of their social relationships.

High-performing females, unlike males who perform well, appear to rely upon peers for assistance in achievement-related activity, but to be less apt to conform when their personal achievement is not being evaluated. While the latter finding could indicate that high-performing females are more socially aloof and rebellious, if this were the case their conformity should be low under both incentive conditions. It therefore seems more reasonable to attribute females' lack of conformity under affiliation-incentive conditions to their relatively greater confidence of being socially accepted.

The above interpretations, if correct, may clarify substantially the interrelationship of personal achievement and social goal attainment among males and females suggested in the study by Wyer and Terrell (1965). Females who perform well academically may seek assistance from other students and presumably offer assistance in return. In the process of cooperative pursuit of achievement goals, their confidence in being accepted by their peers may increase, and therefore their tendency to conform merely to avoid deviating from group norms may be diminished. On the other hand, males who perform well academically may do so without relying upon others for assistance. Consequently, they do not have the opportunity to develop and maintain effective social relationships with peers, and are more apt to conform in order to increase the likelihood of being socially accepted.

The reason for the difference in social orientation between academically successful males and females is not completely clear. It may be due in part to the fact that active pursuit of academic goals is socially acceptable for females but tends to be unacceptable for college males. Some evidence that "studiousness" is considered more acceptable by females than by males was reported by Wyer and Terrell (1965). The interaction of aptitude and sex, which indicates that high-aptitude males are fairly high in conformity, while high-aptitude females are low in conformity, is also relevant to this question. Since aptitude differences are more likely to be antecedents than consequents of differences in tendencies to conform, these data support the interpretation that manifest competence in academic-achievement-related activity leads to uncertainty in males about their peer-group status, but increases females' confidence in their social acceptance.

Although it seems most justifiable to interpret conformity under affiliation conditions within the framework of the earlier study by Wyer (1966), an alternative interpretation may be worth considering. That is, conformity under these conditions may indicate a general sensitivity to social evaluation which is also manifested in compliance with social role expectations. These expectations are for personal achievement in the case of males but for pursuit of nonachievement goals in the case of females. Among males, both conformity and high academic performance may indicate a tendency to comply with social role expectations, while, among females, high performance and low conformity may indicate a general insensitivity to these expectations. Although results involving academic aptitude would be difficult to explain, this interpretation should not be completely discounted.

The interaction of performance and aptitude indicated that underachievers conformed less than either high performers of high ability or low performers of low ability. While this could indicate that underachievers are generally rebellious and resistant to social influence, it should be noted in Table 1 that the conformity of this group was particularly low only under achievement-incentive conditions. It may be more justifiable to conclude that these students simply place little value upon appearing competent in achievement-related activity, and therefore have little interest in conforming in order to improve their performance.

As hypothesized, low performers of low ability were high in conformity. While it was more pronounced when personal achievement was ostensibly being evaluated, their tendency to conform apparently generalized across both incentive conditions. These students may not only believe that they are less competent than their peers in achievement situations, but because of their failure to achieve personally they may depend more upon their relationships with others for support, and therefore may be more apt to conform to peer-group norms to increase their likelihood of being socially accepted.

While high performers of low measured ability were expected to attach importance to competence in achievement activity, predictions were not made as to whether or not they would rely upon others for success in this activity. The interaction of performance and aptitude indicates that these students were relatively low in conformity under all conditions; this suggests that overachievers are generally socially independent both in achievement-related activity and in situations where achievement is not a factor.

One qualification should be made in conclusions drawn from the results of this study. In all cases, the source of influence was of the same sex as the subject. Several studies suggest that conformity is substantially affected by the sex of the source as well as by the sex of the subject. For example, subjects of both sexes may judge males to be more competent than females in achievement tasks (Bennett & Cohen, 1959; Rosenblith, 1959), and therefore may be more susceptible to informal influence from male sources than from female sources. Furthermore, Weiss (1961) found that females, when engaging in achievement-related activity, attempt to appear less competent than males but not other females. It is conceivable that males who are concerned about their social acceptance will conform to other males, but will try to appear dominant and independent in the presence of females; alternatively, females may be more susceptible to the normative influence of males than of other females. Although the extent to which these factors would affect the results reported in this study is unclear, this possiblity should be kept in mind.

REFERENCES

ASCH, S. E. Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men.* Pittsburgh: Carnegie Press, 1951. Pp. 177–190.

BECKER, S. W., LERNER, M. J., & CARROLL, J. Conformity as a function of birth order, payoff, and type of group pressure. *Journal of Abnormal and Social Psychology,* 1964, **69**, 318–323.

BENNETT, E. M., & COHEN, L. R. Men and women: Personality patterns and contrasts. *Genetic Psychology Monographs,* 1959, **59**, 101–155.

BURGESS, E. Personality factors of over- and underachievers in engineering. *Journal of Educational Psychology,* 1956, **47**, 89–99.

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology,* 1955, **51**, 629–636.

DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *Journal of Abnormal and Social Psychology,* 1956, 53, 100–107.

GILMORE, J. V. A new venture in the testing of motivation. *College Board Review,* 1951, **15**, 221–226.

HARVEY, O. J., & CONSALVI, C. Status and conformity to pressure in informal groups. *Journal of Abnormal and Social Psychology,* 1960, **60**, 182–187.

KELMAN, H. C. Processes of opinion change. *Public Opinion Quarterly,* 1961, **25**, 757–781.

MERRILL, R. M., & MURPHY, D. T. Personality factors and academic achievement in college. *Journal of Counseling Psychology,* 1959, **6**, 207–210.

ROSENBLITH, J. F. Learning by imitation in kindergarten children. *Child Development,* 1959, **30**, 69–80.

WEIGAND, G. Goal aspiration and academic success. *Personnel and Guidance Journal,* 1953, **31**, 458–461.

WEISS, P. Some aspects of femininity. Unpublished doctoral dissertation, University of Colorado, 1961.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

WYER, R. S. Self-acceptance, discrepancy between parents' perceptions of their children, and goal-seeking effectiveness. *Journal of Personality and Social Psychology,* 1965, **2**, 311–316.

WYER, R. S. Effects of incentive to perform well, group attraction, and group acceptance on conformity in a judgmental task. *Journal of Personality and Social Psychology,* 1966, **4**, 21–26.

WYER, R. S., & TERRELL, G. Social role and academic achievement. *Journal of Personality and Social Psychology,* 1965, **2**, 118–121.

WYER, R. S., WEATHERLEY, D., & TERRELL, G. Social role, aggression, and academic achievement. *Journal of Personality and Social Psychology,* 1965, **1**, 645–649.

(Received June 20, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 264-278

CONCEPTUAL COMPLEXITY AND REMOTE-ASSOCIATIVE PROFICIENCY AS CREATIVITY VARIABLES IN A COMPLEX PROBLEM-SOLVING TASK[1]

MARVIN KARLINS[2]

Princeton University

This investigation is concerned with the analysis of 2 theories of creativity relevant to the study of complex human problem solving: Mednick's associative viewpoint and Schroder's conceptual systems approach. The 2 positions are defined and compared to see whether they share a homologous conception of the creative process. The balance of the study reports the results of task performance by individuals varying in creativity (operationally defined by scores received on Mednick and Schroder's creativity measures). It was found that the 2 measures of creativity relate differentially to measures of information-seeking behavior in a complex problem-solving task.

Psychologists have long been concerned with the prediction and understanding of complex problem solving in individuals and groups. One approach to this topic focuses attention on the ways individuals explore the dimensions of a hitherto unknown environment or problem. This concern with an individual's pattern of *information search* represents one possible way of viewing problem-solving behavior. It is a way that deemphasizes the end product of problem solving (the answer) as the important unit of study. It concentrates instead on the earlier phases of problem solving: the ways individuals probe for information and the types of information they seek in solving the problem. The question becomes not how much information is known, but how and what information is acquired in dealing with the unknown. This investigation examines problem solving within this information-acquisition framework and tests the comparative effectiveness of the operations utilized in two theories of creativity relevant to the study of complex human problem solving: Mednick's associative viewpoint and Schroder's conceptual systems approach.

Conceptual Systems Theory[3]

A theoretical framework which views individuals and groups, in interaction with their environment, as active information-processing systems is being developed by Schroder and his associates (Harvey, Hunt, & Schroder, 1961; Schroder, Driver, & Streufert, 1966). Known as conceptual systems theory, this social psychological approach to the study of creativity is distinguishable by its insistence on conjoint consideration of situational and dispositional factors, its S-O-R orientation, and its emphasis on *cognitive structure* as the important organismic variable in behavioral analysis.

There are several characteristics that describe the functioning of an individual's cognitive structure, the most important of which is the *structural complexity* of the system. Over a given range of stimuli, information-processing ability varies among individuals and is measured in terms of its *integrative complexity*. Higher integrative complexity refers to a greater number of perceptual categories for receiving information about the world, and more conceptual or combinatory rules for organizing such units of information. Structural complexity is described as varying along a continuum that represents gradations in integrative complexity. This complexity is domain specific; that is, a person can be integratively complex in one area of his "life space" (e.g., political behavior) and simple in another (e.g., religious behavior). A person's level of structural complexity (conceptual level) is a function of the type of training he received in a given domain. This relationship is explicated in detail in the theoretical statement of Schroder et al., 1966.

---

[1] This research was supported by a 2-year predoctoral fellowship award from the National Institute of Mental Health and Peace Corps Contract PC-72-528.

[2] The author gratefully acknowledges the generous assistance provided by Harold Schroder and Bertram Koslin in all phases of this research undertaking. Thanks are also due John Kennedy and Nathan Kogan for their reading of, and commentary on, the manuscript; Sheldon Blackman and Richard Pargament for their assistance in the statistical analysis of the experimental data; and the scholars at Yale responsible for the Human Relations Area Files.

[3] A summary of conceptual systems literature is available elsewhere (Karlins, 1966).

In conceptual systems theory higher levels of integrative complexity provide the necessary prerequisite for creative behavior. Schroder and his associates are quite specific on emphasizing the necessary role of a well-developed integratory mechanism (high conceptual level) in the creative process:

> When personality structure is taken as the anchor for viewing behavior, then we are focusing upon creativity—the ability to generate diversity and conflict, to evolve alternate organizations or integrations of diverse perceptions and decisions [Schroder et al., 1966, p. 11].

In conceptual systems theory, the Paragraph Completion Inventory (Schroder & Streufert, 1962; Schroder et al., 1966) is the most popular test for assessing integrative complexity (creativity). In the Paragraph Completion Inventory the subject is required to write three-sentence responses to each of five sentence stems, each presented for 90 seconds.

> It will be noted that these items represent the presentation of discrepancy, uncertainty, control or constraint and, therefore, tend to engage the subject in some form of "resolution." The items are also selected to be particularly relevant for interpersonal behavior in social and group task oriented activities [Schroder et al., 1966, p. 190].

Associative Theory[4]

A conception of human behavior which explains creative performance within an associative framework is being developed by Mednick and his colleagues (Mednick, 1962; Mednick & Mednick, 1964). In the major statement of the theory, creativity is defined as the

> . . . forming of associative elements into new combinations which either meet specified requirements or are in some way useful . . . the more mutually remote the elements of the new combination, the more creative the process or solution [Mednick, 1962, p. 221].

According to these investigators, there are three associative mechanisms by which the creative process may occur: (*a*) serendipity, (*b*) similarity, and (*c*) mediation. These mechanisms refer to the stimulus configuration in an individual's environment at a given time, the properties of the associative elements (the similarity and/or overlap of stimulus elements), and how the individual uses these contiguously occurring requisite associative elements as the necessary antecedent building-block components which make possible the construction of new combinations of associative elements.

The personality orientation of the theory is accentuated in its emphasis on individual differences in creativity. Mednick (1962) believes that

> . . . any ability or tendency which serves to bring otherwise mutually remote ideas into contiguity will facilitate a creative solution; any ability or tendency which serves to keep remote ideas from contiguous evocation will inhibit the creative solution [p. 222].

Some of these "abilities or tendencies" that make for individual differences are the: (*a*) number of requisite associative elements available to the individual for combinatory purposes, (*b*) organization of an individual's associative elements in an associative hierarchy (e.g., steep versus level associative hierarchy for response in a given stimulus domain), (*c*) number of associations available to the individual in a given stimulus domain (once a requisite associative element is available in an individual's response repertoire).

The operational definition of creativity takes the form of a test (Remote Associates Test—RAT), the structure of which was suggested by theoretical considerations. A test item consists of three stimulus words which the examinee is asked to relate to each other by means of a single common association (a fourth word that is related by association with the first three). It is important to note that the answer to the three words in the item is associational, and *not* definitional, logical, or conceptual in nature. There are 30 such items in the 40-minute test. Some typical RAT items are COOKIES, SIXTEEN, HEART, ———.

---

[4] A summary of (remote) associative creativity literature is available elsewhere (Karlins, 1966).

The correct answer that should appear in the blank space is "sweet." Cookies are sweet; sweet is part of the phrase "sweet sixteen," and part of the word "sweetheart." As can be seen, the fourth word may be related to the other three in various ways for various reasons.

Theoretical and Operational Statements of Conceptual Systems and Associative Theory Compared

In examining similarities and differences in the two theories of creativity two striking parallels become immediately evident: both viewpoints are concerned with a structural approach to behavior (e.g., *how* not *what* people think) and share two explanatory mechanisms, differentiation and integration (the number of units of information an individual has in a given problem domain and how he combines those units in information processing and decision making), in accounting for individual differences in creative behavior. It is vitally necessary to examine how the two systems utilize these concepts, as they form the foundation of each theory, and are utilized in the construction of theory-relevant hypotheses.

In conceptual systems theory, differentiation refers to the number of dimensions of information available to an individual in a given stimulus domain and the fineness of discrimination along a given dimension attainable by the subject; integration refers to the capacity of an individual to generate combinations of dimensions and interrelate these in a number of ways in information processing and problem solving. High conceptual level individuals, in comparison to their low conceptual level counterparts, interrelate more dimensions in the ordering of stimuli via the utilization of more alternate but related rules for combining these dimensions in information processing. Integratively complex individuals have a greater number of perceptual categories for receiving information about the world and more alternate conceptual or combinatory perspectives and rules for organizing such units of information.

In Mednick's associative theory, differentiation refers to the number of requisite associative elements that are available to the subject for information processing, while integration refers to the organization of an individual's associative elements in an associative hierarchy which varies in slope (availability and response probability of elements to a given stimulus). In this system the highly creative individual, in comparison to his low-creative counterpart, has a greater number of associations to associative elements at his disposal and possesses more associative elements of even associative response strength in his associative hierarchy.

The similarity of the two theories regarding the conceptualization of "differentiation" and "integration" can be summarized as follows: (*a*) Both emphasize a concept of differentiation—conceptual systems in its discussion of elementary dimensions available to the subject, associative theory in its emphasis on requisite associative elements necessary to the creative process; (*b*) both recognize differentiation as a necessary but not sufficient aspect of creativity; (*c*) both recognize a concept of "integration" (e.g., "combination") which is: separate from differentiation; as important, if not more so, than the process of differentiation in the dynamics of the creative process; a step in the creative process that functions by operating upon parts in the system, for example, the unification of two parts into a new product.

With the theoretical similarities outlined above in mind, is it fair to say that the two conceptions share a homologous view of personality functioning? It seems not. *The critical difference between the Mednick and Schroder conception of creativity appears in the consideration of the "integratory" or "combinatory" mechanism in the creative process.* It is this difference that sets the two theories most distinctly apart and, is contended here, indicates that the two viewpoints share a common concept of process of differentiation, but a divergent view concerning the means by which information is "put together" in information processing.

In Mednick's associational framework the combination of stimulus elements takes place through the association of elements with one another. The type of association that is made between any two elements is a function of:

(*a*) the number of associations the subject can bring to bear in a given stimulus domain, and (*b*) the organization of an individual's association in a given stimulus domain (associative hierarchy). It is predicted that the more associations an individual can generate in response to a given stimulus, and the flatter his associative hierarchy around that stimulus, the faster and more probable will be the attainment of a creative solution (e.g., remote associate to that given stimulus).

In Schroder's conceptual systems approach the combination of dimensions (and their various discriminatory scale values) is accomplished through mediating "rules" or "schemata" that determine the way information is processed. The complexity of the schemata which relate the dimensions can vary in terms of the number, interrelatedness, and flexibility of the rules available to the information-processing structure, leading to individual differences in the way people combine units of information.

An examination of the way the two theories view an integratory mechanism raises the following question: can combination in Mednick's system be accomplished by the use of one fixed information-processing rule? That is to say, is a person of low integrative complexity capable of combining associative elements according to Mednick's notion? If he is, of course, then the RAT should not differentiate between integratively simple and complex functioning individuals—as both types of information-processing structures should be able to perform on the RAT equally well. Far more important, highly creative subjects (as defined by their scores on the RAT) cannot be considered highly creative when viewed in a conceptual systems framework—for the very combinatorial system that demarcates creative from noncreative performance in associative theory is accomplished equally well by individuals varying in creativity (as measured by the Paragraph Completion Inventory) by conceptual systems standards.

If, indeed, one fixed conceptual rule will serve to fulfill the combinatory requirements for associative creativity (as operationally defined by performance on the RAT test), then we can say that both conceptually simple and complex individuals may be associatively creative, but that only structurally complex individuals are integratively complex (i.e., creative). What this might suggest is that on a scale running from low to high creative, the associative conception of creativity is distinguishing levels of creativity along a more limited section of the scale than is conceptual systems which postulates higher order combinatory rules that differentiate individuals by level of creativity—individuals who would not be so differentiated by an associative notion alone.

Within a conceptual systems framework, performance on the RAT would be explained in terms of the subject generating dimensions to each test item, which would be compared on the basis of associative commonality through the usage of a simple combinatorial rule. No appeal to higher order integrative mechanisms (i.e., rules for relating and comparing alternate structures and generating alternate organizations of relationships) would be needed in explaining performance in the task.

In summary, then, it seems that although both conceptual systems and associative theory recognize a combinatorial principle in explaining creative behavior, the principles are not the same. The differences between the two conceptions of how information is combined in the creative process is evidenced in the type of combinatorial activity required of subjects in RAT performance. Based on theoretical expectations, this test need not differentiate between individuals varying in conceptual level due to the fact that higher order integratory mechanisms (possessed only by integratively complex individuals) are not necessary to solve RAT items. To the conceptual systems theorist, associative creativity can be explained by appealing to relatively simple and fixed rules of information processing, without the necessity of calling forth more flexible, alternate ways for dealing with the organization of information. Empirically, there is evidence that supports the notion of a divergence between what the two tests are measuring: a recent study (Schroder, Harvey, Hunt, & Koslin, 1965) found a .01 *correlation* between the scores of 187 Peace Corps volunteers on the creativity indexes.

Relation of Two Operational Definitions of Creativity to Measures of Information-Seeking Behavior in a Complex Problem-Solving Task

One of the major problems in understanding the dynamics of creativity results from the dearth of controlled laboratory studies relating the creative process to human problem-solving behavior. It often seems that in the attempt to support specific theoretical contentions in the framework of pencil-and-paper-test correlations or judged creativity one ignores the examination of creativity in other behavioral situations. Jackson and Messick (1965), commenting on the problem of previous creativity research, feel that:

> . . . the controversy created by some of the research studies of creativity seems to have focused attention not on unresolved conceptual problems but on the empirical specification of particular, and sometimes arbitrary, aspects of the creativity-intelligence distinction. Moreover, the interesting results obtained with tests of fluency and remote associations have diverted our attention temporarily from the important task of considering other criteria, in addition to the production of unusual responses, essential to the total assessment of creativity [pp. 309–310].

It is the purpose of the present investigation to study the creative process in an ongoing problem-solving situation in which the subject, through his information search and processing prerogatives, is given maximum degrees of freedom in bringing meaning to an initially unstructured environment. The means by which he accomplishes this task is of primary interest, as it is predicted that the creativity level of the individual problem solver (operationally defined by scores received on Mednick and Schroder's creativity tests) will influence the ways in which he attacks the problem. In the present research the Inductive Teaching Program will be utilized as a means of examining the ways individuals search for information in the initial stages of solving a complex problem: the Community Development Exercise. A description of the Inductive Teaching Program and Community Development Exercise (showing how they provide the framework for testing hypotheses) will be presented in the Method section of this paper.

Experimental Hypotheses

In the Community Development Exercise, an individual is given a limited amount of time to solve a complex problem. Through information search the subject is expected to learn about an unfamiliar South Seas island culture—asking enough questions about the island and its inhabitants to find the *best possible way* to get native approval and cooperation in building a needed hospital. The subject is informed that the more relevant information he can gain concerning his problem, the more likely it will be that he will find the best solution to his problem. The types of information available to the subject are presented in the 57-category fixed program deck of the Inductive Teaching Program. These 57 elements (or dimensions) consist of a word or short phrase and represent types of information that may be requested concerning the task. The kinds of information presented in the fixed program deck are varied and numerous. Further, some of these elements are more obviously related to the problem than others. It is assumed that each time the subject utilizes one of these 57 units of information it is because he has been able to see a relationship between that unit of information and the problem under consideration.

In his information request, where the subject must utilize the fixed program deck items in order to gain information he feels important to solving his task,[5] it is predicted that the creativity level of the individual problem solver will influence the ways in which he searches for information. Behaviorally, these differences will be manifested in two ways: (*a*) the number of information domains the subject employs in dealing with the complex task that confronts him (number of fixed program deck categories sampled), (*b*) the frequency with which each information domain is used in information request (number of times each fixed program deck category is sampled). The first way refers to the *breadth of information search* (across categories); the second, to the *evenness*, or *breadth and depth*,

---

[5] Each time the subject wished to ask a question, he was instructed to place the question in the most appropriate category in the fixed program deck (he wrote the fixed program deck category on his question card). In this way, the experimenter was provided with a quantitative means of examining breadth and evenness of category search, based on the subject's own decisions concerning category placement.

of information search (relative frequency of sampling within and between categories). Breadth of information search is measured by determining the number of fixed program deck categories sampled (with analysis of covariance controlling for possible bias due to differences in questioning frequency); breadth and depth of information search is measured through utilization of the *H* (uncertainty) statistic (Senders, 1958).

The way these measures will be utilized to examine experimental hypotheses generated by Mednick's and Schroder's theories of creativity will now be outlined, along with the rationale underlying the hypotheses.

1. It is hypothesized that integratively complex subjects, in comparison to integratively simple subjects, will exhibit more breadth of category search (with a correction for questioning frequency), as reflected in a sampling of more fixed program deck categories during 135 minutes of information request.

This hypothesis is in keeping with previous studies which found that individuals with a high integrative complexity index (compared to those with lower index scores) request more different kinds of information in *complex*[6] problem-solving tasks (Dold, 1964; Driver, 1960; Karlins, 1965; Streufert & Schroder, 1965; Tuckman, 1964). These studies, and the experimental hypothesis, say in effect that conceptually complex individuals are more diverse in their information search when that information search requires the generation of perspectives, the interrelating of two or more such perspectives, and the ability to generate increasingly higher order rules for dealing with information. In the present experiment it is expected that integratively complex individuals (in comparison to integratively simple individuals) will sample a greater number of fixed program deck categories because they are able to generate perspectives that link these categories with task considerations—an ability that requires higher order combinatorial rules[7] of information processing.

2. It is hypothesized that integratively complex subjects, in comparison to integratively simple subjects, will exhibit more evenness in their information search, as reflected in a more equal sampling in and among a greater number of fixed program deck categories (statistically, a higher uncertainty score).

In the framework of the present hypothesis the question now becomes: once the subject has a perspective that permits him to explore a given category in relation to his task, what type of exploration takes place? Does he generate one question in a given category and then turn his attention elsewhere? Does he generate all his questions in a few categories, to the exclusion of the rest? Does he ask a few questions in many categories?

It is possible for individuals to sample the same number of categories (identical breadth of information search) asking the same number of questions and still have different *patterns* of information search (evenness of category sampling). Two subjects, A and B, might both ask 100 questions in 10 different categories, but their evenness of information search would differ if Subject A asked 90% of his questions in one category while Subject B asked 10% of his questions in each category.

In terms of the present uncertainty analysis one can say that Individual B, in comparison to Individual A: (*a*) displayed more evenness of information search across and within categories, (*b*) generated more uncertainty, in the statistical sense (the number of binary questions that must be asked in order to ascertain his behavior).

It is predicted that integratively complex individuals (in comparison with integratively simple subjects) will sample more evenly across and within the fixed program deck categories, based on their ability to tolerate uncertainty (e.g., not spend time exhausting the information in a few categories, but search evenly over many), and become involved with a greater number of categories due to more complex rules which allow for the comparison and linkage of different perspectives to task considerations.

---

[6] The necessity of providing a *complex* problem-solving task is based on theoretical considerations. It is a major contention of conceptual systems thinking that the level of environmental complexity influences the degree to which individual *performance* differences in the acquisition of information can be manifested in a given problem-solving situation. One would *not* predict information-search differences between subjects varying in integrative complexity in simple-problem environments that do not require subjects to utilize higher order combinatorial rules in dealing with the problem.

[7] The process of differentiation should not be a factor here since all subjects are given the 57 categories at the outset of the task.

Breadth and Evenness of Information-Search Hypotheses Applied to Associative Creativity

It is assumed that each time the subject utilizes one of the 57 units of information from the fixed program deck it is because he has been able to see a relationship (make an association) between the unit of information and the problem under consideration. In terms of associative theory, it is assumed that the subject can reach a creative solution utilizing the 57 elements, through the process of serendipity and/or mediation. It is assumed further that the 57 fixed program deck items are analogous to word responses of different associative response strength in a given associative hierarchy. That is, some of the 57 categories in the present task are more "remote" to the problem than others; just as some words (e.g., "Mabel," "food") are more "remote" to the word "table" than others (e.g., "chair," "cloth").

Figures 1 and 2 present this comparison pictorially. As can be seen, there is a goodness of fit when the probability of utilization of fixed program deck items is considered within Mednick's associative hierarchy framework. The relationship between the two figures is very close: associative response strength is plotted on both ordinates and verbal elements along both abscissas. Both theoretical models predict that the number and order of responses elicited to a given stimulus or problem is a function of the associative hierarchy around that stimulus or problem. That is, responses to a given problem (stimulus) vary in their probability and order of evocation, based on their associative response strength in the associative hierarchy. In both models, the more creative individuals would be identified with a flat associative hierarchy, as this would indicate a greater number of associations with a likelihood of response evocation to a given problem or stimulus. This means that in a given problem area, the flatter the gradient of associative response strength across possible associative responses in a given associative hierarchy, the more likely the possibility that an individual will utilize more of the responses in dealing with the problem.

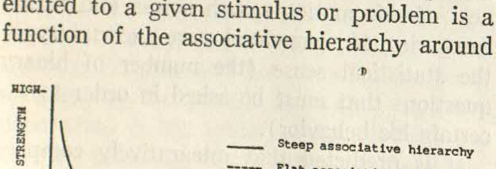


FIG. 1. Associative hierarchies around the word "table."

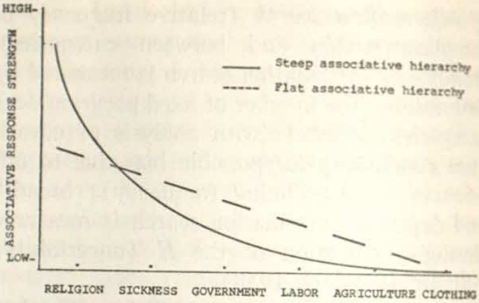


FIG. 2. Associative hierarchies around the problem: "Build a hospital with the aid and cooperation of the natives."

In both models it is assumed that each time the subject utilizes an element from an associative hierarchy around a given stimulus he has formed a relationship between that component of the hierarchy and the stimulus around which the hierarchy is based. The theoretical description of such a relationship is "the making of an association"; the behavioral manifestation of such a relationship is the *elicited* or *utilized* response to a given stimulus or problem. In the case of Figure 1 this behavioral manifestation takes the form of a spoken word (e.g., "chair"); in Figure 2 it is manifested in utilization of an element (e.g., fixed program deck category).

With these considerations in mind, the following hypotheses, and the rationale underlying them, may be stated:

3. It is hypothesized that high associatively creative individuals, in comparison to low associatively creative individuals, will exhibit more breadth of category search (with a cor-

rection for questioning frequency), as reflected in a sampling of more fixed program deck categories during 135 minutes of information request.

Two of Mednick's main operations for creativity are the number of associations that an individual has to the requisite elements of a problem, and the shape of the associative hierarchy in a given problem domain. In this system the highly creative individual (in comparison to his low-creative counterpart) has a greater number of associations to associative elements at his disposal, and possesses more associative elements of even associative response strength in his associative hierarchy in a given problem domain.

Both the above operations lead to the prediction of a greater breadth of information search by highly creative individuals, for the following reasons:

Number of associations. It is assumed that the larger the number of associations an individual has to the requisite elements of the problem (in this case, the number of associations that can be made to each category of the fixed program deck), the higher the likelihood that a relationship between any given fixed program deck category and the problem will be achieved, leading to the use of that category in exploration of the problem.

Shape of the associative hierarchy. As Figure 2 indicates, the highly creative individual, with a more evenly distributed response strength across the 57 categories of the fixed program deck (flat associative hierarchy), will be more likely to utilize more categories than an individual who dissipates his information request over fewer categories (steep associative hierarchy). Mednick (1962, p. 223) points out that an individual whose associative hierarchy is characterized by a flat slope will ". . . be able to get to the less probable, more remote kinds of associations . . ." In this framework the more remote kinds of associations are represented by fixed program deck categories that are less obviously related to the problem solution.

4. It is hypothesized that high associatively creative subjects, in comparison to low associatively creative subjects, will exhibit more evenness in their information search, as reflected in a more equal sampling in and among a greater number of fixed program deck categories (statistically, a higher uncertainty score).

The rationale underlying the previous hypothesis forms the basis of this one, also. Actually, one would expect that breadth and evenness of information search would be related, due to the fact that the more questions an individual expends in searching out a given category, the less the likelihood of questions being asked in a number of categories. Thus, the person who searches over a great many categories is not likely to pile up questions in a few; conversely, a person who searches heavily in a few categories is not likely to search widely over many categories. The interrelationship between these two measures is somewhat analogous to the overlap between Mednick's notion of number of associations and associative hierarchy. In this case the number of associations to a given stimulus should be correlated with the shape of the associative hierarchy with regard to that stimulus since ". . . an individual with a high concentration of associative strength in a few associative responses is not likely to have a proliferation of associations [Mednick, 1962, p. 224]."

5. It is hypothesized that integratively complex individuals, in comparison to integratively simple subjects, will ask *fewer* questions that contain external referents. An external referent is defined as information pertaining to matters not completely concerned with the native culture (e.g., information external to the culture proper). An external referent question is operationally defined as any information-request card containing one or more of the following words: "missionary," "mainland doctor," "stranger," "visitor," "trader," "white man," "mainland government." No predictions are made concerning associative creativity, which is used as a control variable in this hypothesis.

The rationale behind this hypothesis is found in the theoretical contention that high conceptual level individuals have a more well-developed self-concept and sense of internal causation than do low conceptual level individuals. Integratively complex individuals are active manipulators of their world, as they do not perceive rules as externally defined and determined, but internally generated and controlled. To the integratively complex individual, external agents do not provide structure or guide actions; instead, environmental control is viewed in the domain of personal action. Several studies support this notion (Driver, 1962; Harvey, 1963).

In the present problem-solving situation such differences in self-concept should be reflected in the types of questions asked about the native culture. Thus, where it is expected that the integratively complex individual will try to solve his problem by direct intervention in the environment (asking questions about the culture that will aid him in dealing with the island and its inhabitants), this exception does not hold with an integratively simple subject, who will try to solve the problem by appealing to sources that might already have the information he seeks (e.g., the "right" answer).

6. At the outset of the experimental task the subject was required to perform a "levels of importance" card sort. In this card sort the subject was instructed to indicate how *important* the 57 categories of the fixed program deck would be in helping to achieve the goal of constructing the hospital with the cooperation of the natives. This is done by sorting the categories (printed on cards) into groupings which vary in terms of how important the information on the cards will be to the success of the project. It was up to the subject to decide how many of the 57 cards should go into each grouping, and how many groups there would be. The levels of importance card sort was carried out twice during the experiment: before and after the periods of information search.

It is hypothesized that integratively complex subjects, in comparison to integratively simple individuals, will make more differentiations among fixed program deck items in their second levels of importance card sort than their first, as measured by the change in the number of levels constructed between the two tasks. No predictions are made concerning associative creativity, which is used as a control variable in this hypothesis.

In the present investigation the subject performed his first card sort to "familiarize him with the fixed program deck categories to be employed in the information-request periods." The second sort took place after the periods of information request—after the subject had an opportunity to process and evaluate information concerning the problem. It is assumed that exposure to the culture (problem) should have an effect on the subject's perception of the categories dealing with the culture; that is, based on requested and processed information, categories should become more meaningful, giving the subject a better idea of their relative importance to one another. Changes in the number of levels constructed between the two card sorts are expected to reflect changes in perceived importance of information pertaining to the culture, based on experience with the culture.

For the integratively complex individual it is expected that experience with the task environment will be reflected in a greater number of differentiations between the before- and after-task card sorts. That is to say, the integratively complex individual, more effectively exploring the problem environment than his integratively simple counterpart, will be able to make finer discriminations among categories in terms of their importance to the problem. This should lead to more differentiations of importance among categories in the second, compared to the first, card sort.

METHOD

Subjects

Sixty Princeton University male students varying concomitantly in conceptual level and associative (remote) proficiency, and matched on intelligence, were selected from an initial sample of 300 subjects and assigned to one of four experimental groups (integratively simple, low-RAT scorers; integratively simple, high-RAT scorers; integratively complex, low-RAT scorers; integratively complex, high-RAT scorers). Subjects were selected on the basis of three test scores: an intelligence inventory (Wonderlic, 1952)[8] and two creativity indexes (Mednick's, 1959, RAT, Form 1; Schroder & Streufert's, 1962, Paragraph Completion Inventory). A correlation between scores on the RAT and Paragraph Completion Inventory (based on the total subject sample of 300 subjects) was low (.07), replicating the earlier finding in the Schroder et al. Peace Corps study.

Table 1 presents pertinent information on the subject sample. The following requirements were met in subject selection: (*a*) *No differences in intelligence scores between the four experimental*

---

[8] The Wonderlic Personnel Test is a 50-item, 12-minute test of general intelligence.

TABLE 1

CHARACTERISTICS OF THE SUBJECT SAMPLE

Test	*M*	*SD*
Wonderlic Personnel Inventory	39.21	5.89
Paragraph Completion Inventory	5.26	2.10
RAT	20.02	4.86

Experimental groups	Group means for the 3 tests of selection		
	Wonderlic	RAT	Paragraph Completion Inventory
Integratively simple, low-RAT scorers	40.66	14.53	2.90
Integratively simple, high-RAT scorers	40.73	24.60	2.88
Integratively complex, low-RAT scorers	40.06	15.60	8.45
Integratively complex, high-RAT scorers	40.73	25.53	8.11

Note.—Total subject sample information ($N = 300$). Selected subject sample information ($N = 60$, 15 per experimental group).

groups. (*b*) *Differences in remote associates scores between the low- and high-creative groups.* No subject was selected as a high or low creative unless his RAT score fell in the lower or upper one-third of the total distribution of RAT scores based on the original sample of 300 subjects. The *average* score of the low-RAT subjects used in this research fell at the thirty-fifth percentile of a sample of 2,786 University of Michigan students; the *average* score of the high-RAT subjects fell at the ninety-fifth percentile of the Michigan sample. (*c*) *Differences in paragraph completion scores between the low- and high-creative groups.* In this study, the mean integrative complexity score of the integratively simple sample was 2.89, indicating a general conceptual orientation characterized by the ability to generate differentiation along a dimension in the domains of uncertainty and conflict (as expressed in the Paragraph Completion Inventory), and the emerging tendency to utilize an alternate perspective. The mean integrative complexity score of the integratively complex sample used in the study was 8.28. At this level alternative perspectives (multidimensional) characterize information processing, and poorly formed interrelating linkages between perspectives are in evidence. Consequently, in this sample, we are comparing groups which are at the level of well-articulated single-perspective conceptualization and loosely related organization of multiperspectives (combinatory conceptual rules).

Experimental Task: The Inductive Teaching Program

A full description of the Inductive Teaching Program is presented elsewhere (Karlins & Schroder, 1967), along with the philosophy underlying the device, and the interested reader is referred to this source. Briefly, the Inductive Teaching Program contains a set of facts about a specific problem or situation, stored on individual cards, analogous to the memory storage banks utilized by computers. The subject is charged with the responsibility of getting these facts by the "inquiry" method, asking questions that will solicit these pieces of information. Keeping in the computer analogy, the subject writes a program to have the computer retrieve a certain stored bit of information. Each fact is stored on an IBM punch card—typewritten on the blank side if the information is longer than 80 alphabetic characters, punched and interpreted (printed) on the front side if the information is 80 characters or less. The idea of developing memory storage banks with enough information units to answer almost any question posed by the subject in a given situation has already been accomplished in a tactical war game utilizing a fixed-input program (Karlins, Schroder, & Streufert, 1965), in a navy task-force simulation (Karlins, Coffman, Lamm, & Schroder, 1967), and the Community Development Exercise reported in this research.

In the Inductive Teaching Program the subject is active in every phase of the learning process—from his active information search and processing of requested information to his utilization of this information in decision making. Inductive teaching is similar to programmed learning in the sense that the subject is allowed to control the rate at which information is received; it is different than programmed learning in that it makes the subject active in the information acquisition and organization process. The Inductive Teaching Program is similar in many respects to the autotelic environment of Anderson and Moore (1960), as it provides the subject with a rich environment to explore and manipulate, receiving feedback relevant to, and based upon, his behavior. The Inductive Teaching Program can be summarized by: its flexible teaching pattern—the subject determines what will be taught (and when) by his pattern of information request; flexible, responsive environment that provides feedback to the subject relevant to his inputs; demands on the subject for active manipulation of information; emphasis on interactive connections between the subject and the environment he is investigating; storage of *factual* material only.

The most important aspect of the Inductive Teaching Program is its emphasis on decision making through subject-instigated inductive reasoning: the Inductive Teaching Program provides the subject with the basic units of information necessary for decision making, while requiring him to accumulate and process the information in self-meaningful and relevant ways for problem solving. This is indeed the most important underlying consideration of the Inductive Teaching Program: the informational program deck (deck of answer cards to the subject's questions) serves as a repository of facts, not a storehouse of synthesized or creative problem solutions. In his questioning the subject may, through information search, receive facts concerning the problem he is interested in, *but all assumptions, inferences, and extrapolations from these facts must be generated by the subject*. If Newton were learning via the Inductive Teaching Program, he could receive the piece of information that "apples fall from the tree branches to the ground below," but any postulation of a law of gravity from this information would be an induction and thus, of necessity, be generated by Newton, not the Inductive Teaching Program which offers no such information.

An example might bring the philosophy behind the Inductive Teaching Program into sharper focus. In the research reported in this investigation the subjects were required to find the best possible way to build a hospital, with the cooperation and aid of local natives, on the South Seas island of "Wabowa." As part of the solution to this problem the subject might want to know the best location for building the hospital. There are two ways he could attempt to find an answer to his question: (*a*) He could send in a question card asking, "Where is the best place to build a hospital on Wabowa?"; or (*b*) he might ask several questions related to the location problem (e.g., type, availability, and distribution of natural resources, terrain of the land, transportation facilities, population concentrations, labor supply, climatic conditions, availability of land, etc.) that, when put together, would allow him to make a decision as to the best construction site. In the former case the subject has asked an "inferential question" that requires an inferential problem solution not available to him through the factual information contained in the Inductive Teaching Program. In the latter case the subject has made the proper sequence of inquiries: he has asked for factual information which can be used to arrive at the answer he seeks.

Composition of the Inductive Teaching Program. The Inductive Teaching Program normally consists of four interrelated component parts:

1. The fixed program deck. The fixed program deck consists of information about a given environment transmitted to the subject without his request. Such items can take the form of informational items representative of a larger body of material the subject must eventually ask questions about or categories (topics) relating to various aspects of a given environment. If such categories are supplied to the subject, he is required to utilize them in his question request, for example, assign each question he asks to the most appropriate category available.

2. The reactive program deck. This program contains all the answer cards available on a given environment (problem), one or more of which is presented to the subject when he asks specific questions concerning the task he is exploring. The number and type of reactive program deck cards given to the subjects are entirely determined by the number and type of questions the subject asks. The reactive program deck gets its name from the fact that, unlike the fixed program deck items which are given to the subject regardless of his actions, the reactive program deck "reacts" to the subject's information requests with appropriate answers.

3. Rule violation cards. The subject is informed (in task instructions) of five writing restrictions that are to be observed in question composition. Rule violation cards replace answer cards when more than one question is asked per card, questions concern information already discussed in task instructions, questions are inferential in nature (require 'inferential" rather than factual answers to question), an answer to a question has already been received on a previous card, questions are longer than two sentences. Rule violation cards serve the important function of imposing necessary restrictions on the subject's compositional output, insuring necessary uniformity in the form, length, and type of questions asked.

4. No answer cards. In his task instructions the subject is informed that

> . . . we believe that we have answered the majority of any questions that might arise, however, some information might not be available concerning your problem, either because that aspect of the problem has not been studied or because we have not anticipated your question. In the case where your question cannot be answered you will receive a card stating that no information is available on the question asked.

Inductive Teaching Program applied to the present research: The Community Development Exercise. In the Community Development Exercise each subject was asked to imagine that he was a Peace Corps volunteer who had just received his overseas assignment. His supervisor has just informed him that a mainland government in the South Pacific has requested his aid in a community-development project on the island of Wabowa, some 250 miles from its coast. His assignment is to build, in a cooperative effort with the local natives, a hospital on the island. He was informed that a similar task had been attempted by missionaries 3 years before, but that these individuals were unable to enlist native aid, and the project failed. The importance of getting the hospital built was emphasized to each subject: venereal disease was spreading rapidly on the island—killing the natives and infecting traders in the area—much to the chagrin of the mainland government. Utilizing his information-search prerogatives, the subject was expected to learn about the culture of the islanders he would be living with, and in such a way present a tentative solution to the problem, that is, his estimation of the *best possible way* to get native approval and cooperation in building the hospital.

The major problem confronted in the Community Development Exercise was the design of the Wabowan culture. This was accomplished by basing the Wabowan culture on the Trobriand islanders, a South Seas tribe studied extensively by Malinowski and described in his book *Argonauts of the Western Pacific* (1922). Malinowski's scholarship was instrumental in implementing research goals, for by presenting a detailed, elaborate, and exacting picture of

Trobrianders, set in the larger cultural matrix of the interacting island tribes of the Trobriand archipelago, he presented a culture with enough richness to serve as a useful model for generating the necessary complexity in the Wabowan culture.

The design of the Inductive Teaching Program for the Wabowan culture was accomplished in two stages, the pilot study (Karlins, 1965; Karlins & Lamm, 1965) and the present research. In the first stage of design 1,250 units of information on the Wabowan islanders were composed and placed into 57 categories. Each of the 57 categories contained between 10 and 30 pieces of information. To make answer length comparable, and to aid in cheap, accurate, and rapid reproduction of these answers, all finalized items of information were punched onto IBM cards. Each item of information was expressed in a sentence between 40 and 65 characters in length. Each of these items was then reproduced 50 times, using the 1410 IBM computer, and interpreted (printed). Over 30,000 cards were prepared in this manner. An experiment was then undertaken with 187 Peace Corps volunteers as subjects. This subject sample generated over 10,000 questions concerning the task problem.

The second stage of design consisted of utilizing the 10,000 questions to improve the fixed and reactive decks in the following ways:

1. A new list of 57 categories for the fixed program deck was constructed, making the categories more inclusive, more understandable, and more mutually exclusive in relation to other categories. The effectiveness of the fixed program deck was greatly enhanced by extensive utilization of the coding and categorizing techniques developed by the Human Relations Area Files (Murdock, Ford, Hudson, Kennedy, Simmons, & Whiting, 1950).

2. Answers were prepared to questions that were unanswerable with the pilot study reactive program deck.

3. Some answers were expanded beyond one sentence in length to facilitate better understanding of the item while preserving the richness of the answer.

4. A more effective retrieval system for the reactive program deck was devised. In its final form the second-stage reactive program deck was more than twice the size of its first-stage counterpart (2,755 items) and was far more capable of answering the majority of questions asked by the subjects —a requirement that is vital to the proper functioning of the Inductive Teaching Program.

Procedure

Each subject was run individually by the same experimenter. The experimenter did not know the subject's experimental group classification. The experimenter observed the subject and transmitted written answers to his questions from the other side of a one-way mirror. Questions and answers were received and delivered via mail chutes. In all phases of this experiment there was no direct contact between the experimenter and the subject (except for the reading of instructions over a public-address system). In the experiment the subject performed two levels of importance card sorts (30 minutes each), one before and one after the information-request segment of the investigation (three 45-minute question-asking sessions). The total experiment was completed in one testing session (including several rest periods). Subjects did not mind the length of the experimental session. They were paid $10 for their participation in the investigation.

RESULTS

Six hypotheses were tested in the present investigation: breadth of information search (Hypotheses 1 and 3), evenness of information search (Hypotheses 2 and 4), utilization of external referents (Hypothesis 5), and constructed levels in importance card sorts (Hypothesis 6). Associative creativity was used as an experimental variable in Hypotheses 3 and 4 and as a control variable in Hypotheses 5 and 6. Conceptual systems was used as an experimental variable in all but two hypotheses (3 and 4). Hypotheses 2, 4, and 5 were analyzed using a 2 × 2 factorial design. Hypotheses 1, 3, and 6 were examined utilizing analysis of covariance.

Before turning to the results for each specific hypothesis, a general statement can be made concerning the findings: of the four hypotheses relevant to conceptual systems theory, all were significant at the .05 level of significance (Hypotheses 1, 2, 5, and 6), although the data relevant to Hypothesis 6 is

TABLE 2

RESULTS OF ANALYSIS OF COVARIANCE (CONTROLLING FOR NUMBER OF QUESTIONS ASKED) BEARING ON BREADTH OF INFORMATION-SEARCH HYPOTHESES

Source	SS'	MSS'	F
Systems (A)	335.40	335.40	11.40**
Associative (B)	25.28	25.28	<1
A × B	22.59	22.59	<1
Error	1,618.91	29.43	

Adjusted group means[a]

		Associative	
		Low	High
Systems	Low	31.47	34.03
	High	37.55	37.62

[a] Special formula for computing analysis of covariance adjusted cell means is found in Winer (1962, p. 598).

** $p < .005$, $df = 1/55$.

TABLE 3

RESULTS OF ANALYSIS OF VARIANCE BEARING ON EVENNESS OF INFORMATION-SEARCH HYPOTHESES

Source	*MSS*	*F*
Systems (A)	0.70	5.03*
Associative (B)	0.32	2.31
A × B	0.00	0.03
Error	0.14	

Group means

		Associative Low	Associative High
Systems	Low	4.41	4.57
	High	4.64	4.77

* $p < .05$, $df = 1/56$.

only partially supportive of that hypothesis. In no instance (either as an experimental or control variable) did associative creativity account for significant subject performance differences in any of the experimental tasks.

Table 2 presents information relevant to the two breadth of information search hypotheses. The prediction that integratively complex subjects, in comparison to integratively simple subjects, would sample more fixed program deck categories during 135 minutes of information request was upheld ($F = 11.40$, $p < .005$). The same prediction made in relation to individuals varying in associative creativity was not supported. There was no Systems × Creativity interaction effect.

TABLE 4

RESULTS OF ANALYSIS OF VARIANCE BEARING ON THE EXTERNAL REFERENTS HYPOTHESIS

Source	*MSS*	*F*
Systems (A)	91.27	5.36*
Associative (B)	17.07	1.00
A × B	15.00	0.88
Error	17.03	

Group means

		Associative Low	Associative High
Systems	Low	7.80	5.73
	High	4.33	4.27

* $p < .05$, $df = 1/56$.

TABLE 5

RESULTS OF ANALYSIS OF COVARIANCE (CONTROLLING FOR INITIAL NUMBER OF DIFFERENTIATIONS) BEARING ON CARD-SORT LEVELS HYPOTHESIS

Source	SS_x	SS_y	$SS_{y'}$	*MSS'*	*F*
Systems (A)	.82	43.35	35.98	35.98	4.76*
Associative (B)	2.82	.01	.93	.93	< 1
A × B	8.81	2.82	.06	.06	< 1
Error	562.40	654.00	415.55	7.56	

* $p < .05$, $df = 1/55$.

Table 3 summarizes findings concerning the hypotheses that individuals varying in creativity would exhibit more evenness in their information search, as reflected in a more equal sampling in and among a greater number of fixed program deck categories (statistically, a higher uncertainty score). Once again, there was a significant main effect by conceptual systems ($F = 5.03$, $p < .05$), while no difference was found in relation to associative creativity. There were no interaction effects.

Findings bearing on the hypothesis that integratively complex individuals, in comparison to integratively simple subjects, will ask fewer questions that contain external referents are presented in Table 4. The results are in support of the hypothesis at the .05 level of significance ($F = 5.32$, $p < .05$). Associative creativity (utilized as a control variable) and the interaction term were not statistically significant.

Table 5 presents the analysis of covariance figures bearing on the hypothesis that integratively complex individuals, in comparison to integratively simple individuals, would make more differentiations among fixed program deck items in their second (levels of importance) card sort than their first, as measured by the change in the number of

TABLE 6

MEANS AND ADJUSTED MEANS FOR NUMBER OF LEVELS OF IMPORTANCE CONSTRUCTED IN TWO CARD SORTS

	Integratively simple *S*s	Integratively complex *S*s
Mean for the 1st card sort	8.93	9.16
Adjusted mean for the 2nd card sort	7.11	8.80

TABLE 7

RESULTS OF THE ANALYSIS OF VARIANCE BEARING ON THE NUMBER OF LEVELS CONSTRUCTED IN THE FIRST LEVELS OF IMPORTANCE CARD SORT

Source	*df*	*MSS*	*F*
Systems (A)	1	0.82	0.08
Associative (B)	1	2.82	0.28
A × B	1	8.82	0.88
Error	56	10.04	

levels constructed between the two tasks. An *F* of 4.76, significant at the .05 level, was found. A closer examination of the data, however, indicated only partial support for the hypothesis.

The mean number of importance categories (levels) constructed for the first card sort by integratively complex and simple individuals are presented in Table 6 along with the adjusted means (controlling for the initial number of differentiations) for the second sort. A separate analysis of variance (Table 7) indicated that there were no initial differences between integratively complex or simple individuals in the number of categories constructed in the first card sort. The results indicate that the integratively complex individuals do not make more differentiations in levels of importance between their first and second card sorts, while the integratively simple individuals make significantly fewer differentiations in their second (as compared to first) card sort. Thus, conceptually complex subjects do not make increased differentiations between their two levels of importance card sorts, but do make a significantly greater number of differentiations than conceptually simple individuals on the card sort that follows experience with a complex problem-solving environment.

DISCUSSION

The findings of the present investigation support the notion that subjects varying in conceptual level perform differentially in complex problem-solving situations. Integratively complex subjects, in comparison with integratively simple subjects, (*a*) exhibit greater breadth in their information search, as reflected in the number of fixed program deck categories they utilize; (*b*) search more evenly across and within the informational domains provided by the Fixed Program Deck, as ascertained by the amount of statistical uncertainty (H) they generate; (*c*) ask more questions that indicate a willingness to act upon and explore their environment directly, as adjudged by the infrequency of requests for information pertaining to how other individuals dealt with the problem ("external referents"); and (*d*) maintain differentiation among classes of information dealing with a complex problem-solving environment, as reflected in the "levels of importance" constructed in a before- and after-task card sort.

There are no statistically significant behavioral differences between individuals varying in associative creativity in any aspect of the experimental problem studied in this investigation.

A major question growing out of the research findings concerns the reason why individuals varying in associative creativity did not perform differentially on the experimental tasks. One possible explanation is that the present problem did not tap skills possessed differentially by individuals varying in associative creativity. To be sure, the present problem was far more "open" than those normally employed to test for performance differences in subjects varying in associative creativity. The most common experimental means of testing associative creativity in the past has been through verbal behavior studies. One cannot deny that answering a RAT item or solving an anagram problem calls for a more task-specific, well-defined, and structured answer than that required in dealing with a complex problem like "build a hospital with the aid and cooperation of the natives." Indeed, the major difference between traditional laboratory methods of examining performance differences in subjects varying in associative creativity and the one employed in this investigation is the degrees of freedom allowed the subject in dealing with his problem.

Do the findings of this investigation suggest that the remote associative model is an inappropriate way to view creativity? Definitely not. All that can be said on the basis of this research is that two measures of creativity relate differentially to measures of information-seeking behavior in a complex problem-solving task. There is a possibility that the information-search behavior examined in this investigation has little to do with creative behavior. Maybe greater breadth and evenness in information acquisition is not related to the ability to reach a creative problem solution utilizing that accumulated information. On its face it would seem that the propensity to explore more widely in a given environment would be associated with the ability to deal with that environment creatively, but it is the task of future research to provide an empirical basis for such a belief, if such a belief is correct.

REFERENCES

ANDERSON, A. R., & MOORE, O. K. Autotelic folk models. *Sociology Quarterly,* 1960, **1**, 203–216.

DOLD, A. W., III. The effects of group composition and stress on heterogeneous teams in a simulated industrial setting. Unpublished senior thesis, Princeton University, 1964.

DRIVER, M. J. The relationship between abstractness of conceptual functioning and group performance in a complex decision-making environment. Unpublished manuscript, Princeton University, 1960.

DRIVER, M. J. Conceptual structure and group processes in an inter-nation simulation: I. The perception of simulated nations. Technical Report No. 9, Project NR 171-055, 1962, Princeton University Contract Nonr 1858(12), Office of Naval Research.

HARVEY, O. J. Self systems, anomie and self esteem. Technical Report No. 9, 1963, University of Colorado. Contract Nonr 1147(07), Office of Naval Research.

HARVEY, O. J., HUNT, D. E., & SCHRODER, H. M. *Conceptual systems and personality organization.* New York: Wiley, 1961.

JACKSON, P. W., & MESSICK, S. The person, the product, and the response: Conceptual problems in the assessment of creativity. *Journal of Personality,* 1965, **33**, 309–329.

KARLINS, M. An initial description of the community development task presented with findings concerning conceptual complexity in Peace Corps volunteers. Unpublished master's thesis, Princeton University, 1965.

KARLINS, M. Conceptual complexity and remote associative proficiency as creativity variables in a complex problem-solving task. Technical Report No. 19, Project NR 171-055, 1966, Princeton University, Contract Nonr 1858(12), Office of Naval Research.

KARLINS, M., COFFMAN, T., LAMM, H., & SCHRODER, H. M. The effect of conceptual complexity on information search in a complex problem-solving task. *Psychonomic Science,* 1967, **7**, 137–138.

KARLINS, M., & LAMM, H. The community development task: A theoretical validation of measures. Appendix 1, Peace Corps Assessment Program, 1965.

KARLINS, M., & SCHRODER, H. M. Discovery learning, creativity, and the Inductive Teaching Program. *Psychological Reports,* 1967, in press.

KARLINS, M., SCHRODER, H. M., & STREUFERT, S. A fixed input program for a tactical war game. (Unpublished manuscript) Washington, D. C.: American Documentation Institute, Auxiliary Publications Project, Photoduplication Service, Library of Congress, 1965, No. 8621.

MALINOWSKI, B. *Argonauts of the Western Pacific.* New York: Dutton, 1922.

MEDNICK, S. A. The associative basis of the creative process. *Psychological Review,* 1962, **69**, 220–232.

MEDNICK, S. A., & MEDNICK, M. T. An associationistic view of creative thinking. In C. W. Taylor (Ed.), *Widening horizons in creativity.* New York: Wiley, 1964. Pp. 54–68.

MURDOCK, G. P., FORD, C. S., HUDSON, A. E., KENNEDY, R., SIMMONS, L. W., & WHITING, J. W. *Outline of cultural materials* (3rd ed.) New Haven: Human Relations Area Files, 1950.

SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing: Individuals and groups in complex social situations.* New York: Holt, Rinehart & Winston, 1966.

SCHRODER, H. M., HARVEY, O. J., HUNT, D. E., & KOSLIN, B. Component assessment in Peace Corps trainees. Peace Corps Report, 1965, Princeton University, Contract Nonr P.C. 72-528.

SCHRODER, H. M., & STREUFERT, S. The measurement of four systems of personality structure varying in level of abstractness: Sentence completion method. Technical Report No. 11, Project NR 171-055, 1962, Princeton University, Contract Nonr 1858(12), Office of Naval Research.

SENDERS, V. L. *Measurement and statistics.* New York: Oxford University Press, 1958.

STREUFERT, S., & SCHRODER, H. M. Conceptual structure, environmental complexity and task performance. *Journal of Experimental Research in Personality,* 1965, **1**, 132–137.

TUCKMAN, B. W. Personality structure, group composition, and group functioning. *Sociometry,* 1964, 27(4), 469–487.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received June 28, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 279-290

DOGMATISM, PRENEGOTIATION EXPERIENCE, AND SIMULATED GROUP REPRESENTATION AS DETERMINANTS OF DYADIC BEHAVIOR IN A BARGAINING SITUATION[1]

DANIEL DRUCKMAN

Institute for Juvenile Research, Chicago, Illinois

A 2 × 2 × 2 design with 15 replications in each cell was used to assess the relative contribution of personality and situational variables in determining conflict resolution. The task was a simulation of the collective-bargaining process. Assignment to roles was made on the basis of extreme scores on a labor-management attitude scale. Unilateral prenegotiation experience was separated from group loyalty, and a modification of the Dogmatism Scale was used as the personality variable. The results were consistent across 4 measures of conflict resolution. Resistance to compromise was associated with unilateral planning of strategy (as contrasted with bilateral discussion) and with high scores on the modified Dogmatism Scale. Whether bargainers represented groups or themselves did not have a significant effect on compromise behavior. Bilateral study experience resulted in more agreement on the ranked importance of the issues than strategy experience. Differential yielding by issue was related to the perceived defensibility of the starting positions on each issue.

Several investigators in the area of conflict resolution have been concerned with the problem of how people approach one another as individuals or group representatives in a situation where they come together to resolve differences and where compromise is possible. Two orientations, characterized by Gladstone (1962), seem especially appropriate to this type of confrontation. An egoistic orientation is defined by a strong motive to win, evidenced in noncompromise behavior, a perception of the other in terms of stereotypical categories, and the judgment of an outgroup product to be of inferior quality. Blake and Mouton (1962) spelled out a similar syndrome which they called "win-lose" motivation. The experimental production of egoistic motivation was demonstrated by Sherif, Harvey, White, Hood, and Sherif (1961) with children in a camp setting, and Blake and Mouton (1962) with adults in an industrial setting. An integrative orientation is one in which the importance of collaboration for mutual benefit is realized. Compromise as a means of conflict resolution is predicted to result from this orientation.

Few investigators have attempted to specify the variables responsible for the predominance of one or the other orientation or have separated individual style from situational variables in accounting for the behavior observed. Instead, explanations have relied upon either group relatedness or personality viewed largely as mutually exclusive alternatives.

Group Relatedness

Blake and Mouton (1962) and Sherif and Sherif (1965) observed that there has been a tendency among many investigators to attribute motivation to factors within the individual rather than to properties of the group in which he is a member. They called this the "psychodynamic fallacy." Committedness, according to these authors, should account for more variance in intergroup posture than should personality.

Studies on intergroup relations have demonstrated the production of competitive motivation after having teams independently develop a position or group product. Unilateral participation, rather than attachment to the group per se, may have been responsible for such behavior as the inability objectively to assess the value of competing positions and competitive nonyielding. The constraints placed upon a representative due to unilateral position formation have been illuminated by Blake (1959), Mack and Snyder (1957), and Rusk (1955). Pearson (1959) claimed that ambiguity serves as an aid to problem solving. Turk and Lefcowitz (1962), in a recent restatement of a major part of Sumner's (1906) ethnocentrism syndrome, considered acts of assertion on the part of representatives to be more likely under high ingroup solidarity (position consensus).

---

[1] This paper is based on a dissertation submitted to the Graduate School of Northwestern University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author gratefully acknowledges Winfred F. Hill and Marjorie Druckman for their helpful comments and aid in preparing this manuscript.

In the human relations training laboratories created by Blake and Mouton (1961, 1962) members interacted only with their own group on the development of an assigned problem, and representatives were elected to debate the merits of their group's product. Deadlock in agreeing on the best solution was the invariable result. A unilateral definition of the problem was apparent insomuch as, after studying the competitor's position, most participants showed that they knew more about their own solution and saw little that the positions shared in common. Bass and Dunteman (1963), Blake and Mouton (1961), and Ferguson and Kelley (1964) produced evaluative bias under conditions in which laboratory groups worked independently on the creation of one or more products. From the subject's point of view it is probable that group problem-solving effectiveness was at stake.

However, in most of the experimental studies of intergroup relations, unilateral participation in the development of a product was not separated from group membership or group representation (e.g., Bass & Dunteman, 1963; Blake & Mouton, 1962; Sherif et al., 1961). Another example of a possible confounding factor was seen in a study by Campbell (1960). He found that two-person union and management teams had difficulty in finding a compromise acceptable to both teams after having formed independent strategies. Is this rigid behavior a function of prenegotiation experience or group loyalty, or is it due to an interaction between these two variables?

A few investigators were able to examine the effects of one of these variables while holding the other constant. Bass (1963) had individuals negotiate as representatives of abstract organizations rather than laboratory groups. Agreement was reached quickly if negotiators did not form strategies or spell out positions before bargaining. That is, when future negotiators defined the problem from a bilateral perspective, compromise resulted without deadlock. However, when they spelled out their own position before bargaining, a significantly longer time period was necessary before agreement could be reached, and negotiations often ended in deadlock. Vegas, Frye, and Cassens (1964) and Ferguson and Kelley (1964) found group membership to be the important variable. The former study demonstrated no perceptual distortion and early agreement among two individuals on which one produced the best essay when neither represented a group product. The latter investigators found that subjects who were members of laboratory groups but who did not participate in the formation of the group product overvalued both the ingroup and its products as much as members who took part in product formation. The extent of bias was correlated with attraction to the group irrespective of taking part in product formation. However, without an orthogonal design it is not possible to determine the relative amount of variance accounted for by each of these variables. The Bass study yielded no information on the effect of group membership, while the Ferguson and Kelley study was inconclusive with respect to unilateral product formation.

Personality

Whether a representative behaves competitively or in a conciliatory fashion may depend partially on attributes that he brings to the situation, other than those determined by his role. While several investigators argue against psychodynamic explanations, some experimental work points toward personality correlates of bargaining behavior.

Some personality variables which have been shown to be related to cooperative or competitive choices in a Prisoner's Dilemma game include internationalism-isolationism (Lutzker, 1960), cooperative, individualistic, or competitive motivational orientations (Deutsch, 1962), family orientation (Crowne, 1966), and F-scale authoritarianism (Deutsch, 1960). However, Robinson and Wilson[2] failed to replicate the latter finding when two-man teams competed. While they found a significant ingroup-outgroup bias in both competitive choices and personality-trait ratings, this was not related to authoritarian attitudes.

A correspondence between behavior at the interpersonal and intergroup levels was provided in the research of Blake and Mouton (1961). The elected representatives shared certain rated behavioral characteristics that distinguished them from nonrepresentatives. These subjects were rated by all participants as significantly more nonyielding, independent, and resistant to external influences during the ingroup caucus. This is also an accurate description of how they behaved in the intergroup session. Thus the variables of personality type, unilateral planning, and group commitment could not be separated in accounting for the nonyielding behavior of representatives during the intergroup debate.

Among the characteristics of the dogmatic personality delimited by Rokeach (1960) is blind loyalty to a system. He presented experimental evidence on the tendency of closed-minded subjects not to defect from instructions coming from an external authority. There were significant differences between closed-minded and open-minded subjects on defection from the experimenter's instructions even when they made no sense. Jamias (1964) demonstrated that closed-minded respondents were significantly affected by the values of the social systems within which they lived, while open-minded participants were not. Thus another individual difference variable may have relevance for the behavior of a group representative.

The intention of the present study was to investigate the relative contribution of personality and situational variables as determinants of orientation in a non-zero-sum, simulated, labor-management bargaining game. The investigator sought to determine whether an egoistic orientation or nonyielding behavior is a function of commitment to a planned bargaining posture. The behavior of groups that had formulated a strategy was contrasted with the behavior of groups which had not previously decided upon a strategy. Lack of significant differences in behavior between the conditions indicates that the manipulation made no difference, but does not exclude the possibility that nonyielding behavior is a function of group loyalty. In order to determine if this is the case it was necessary to examine bargaining behavior on the part of representatives not coming from laboratory groups. The personality variable is discussed in the design section.

METHOD

Design

The manipulations were arranged into a 2 × 2 × 2 design with 15 replications (9 male dyads and 6 female dyads) in each cell. The first variable represented an attempt to produce a unilateral versus a bilateral definition of the problem by manipulation of prebargaining environments. The second variable attempted to separate the effects of group commitment per se from the constraints imposed by position taking. Finally, to see whether personality made a difference, representatives were chosen on the basis of extreme scores on a modification of the Dogmatism Scale devised by Rokeach (1960). Each session was homogeneous in attitudinal disposition and sex. Across all eight conditions 120 dyads or 240 subjects were run through the simulation. A brief discussion of the basis for role assignments will be followed by specific details on each manipulation.[3]

Union versus company. Participants were assigned to either a union (employee role) or company (employer role) team on the basis of their scores on a 24-item attitude scale. The scale is balanced with an equal number of items worded in the directions of prounion, anticompany, antiunion, and procompany. The split-half reliability, corrected by the Spearman-Brown formula, is .72. This selection criterion was used with the intention of increasing the identification of subjects with the position that they were given. Subjects indicated, on a postnegotiation question, that their assigned role coincided highly with their own general beliefs. The mean across all subjects was 2.16 on a 5-point scale ranging from "completely coincided" to "did not coincide at all." The differences in role identification between males and females, company or employer and union or employee representatives, and high and low dogmatics were nonsignificant.

---

[2] C. Robinson and W. Wilson, "Intergroup Attitudes and Strategies in Non Zero Sum Dilemma Games: II. Selective Bias in both Authoritarians and Nonauthoritarians," unpublished manuscript, University of Hawaii, 1965.

[3] The instructions for each condition, questionnaires, attitude scales, and additional information concerning the procedure are contained in Druckman (1966) and may be obtained through University Microfilms, 313 North First Street, Ann Arbor, Michigan.

Typical of the scale are the following items, worded in each of the four directions:

Anticompany	Procompany
Most violence found at picket lines is instigated by management itself.	Management should have the sole right to determine fringe benefits.
Prounion	Antiunion
Unions struggle to keep existing work rules in order to ensure the health and safety of the worker, not to make unnecessary work or featherbed.	The motives governing the actions of top union officials are prestige and financial gain, and not the welfare of the workers.

Approximately the upper and lower quartiles of the labor-management scale distribution were eligible for participation. However, only those also scoring in the upper or lower quartiles of a modified Dogmatism Scale were used. The scores ranged from 33 to 118, with a mean of 76.09. The mean score of the union and employee representatives was 62.73 ($V = 10.91$),[4] while the mean for the company and employer representatives was 89.47 ($V = 10.91$).

Unilateral position formation versus bilateral discussion. In the strategy condition the two teams (or individuals) were separated into different rooms for approximately 40 minutes. The instructions for the group condition were as follows:

> The next 40 minutes will be spent with your own team. You are to use this time to plan your bargaining strategy. First gain an understanding of the issues. Then you may formulate a list of points defining the rationale or arguments for your team's position, on each issue. You should decide on items that you think each man representing you should remain firm on. This strategy may be prepared so as to guide the negotiators through the bargaining session. It may also include potential concessions, anticipations for trading, setting goals, and the like.

In general, it is felt that these strategy instructions set the stage for the more typical prebargaining environment under which collective bargaining takes place.

Joint study conditions were created with the intent of releasing bargainers from commitment to a rigid position. The instructions for the group condition were as follows:

> The next 40 minutes will be spent in bilateral study. You are to use this time for learning as much as possible about company and union perspectives. Study the issues in order to gain understanding of both points of view as well as areas of greater or lesser agreement between proponents of either. Do not formulate or plan any strategies for bargaining from either position. Do not take a position and argue its merits against someone who might profess to the opposite position. Finally, do not form coalitions with other team members to bargain or debate from a position.
>
> If at any point the discussion seems to break down into an interteam competition, the experimenter will stop it and remind participants of the goal. You should act bipartisan with respect to the issues during this session. There is no need to use any form of propaganda in order to obtain later gains. Feel free to enter the discussion at any point and openly discuss the issues. The emphasis is on informality and there is no one observing you or recording data. Understanding is the goal![5]

Blake and Mouton (1962) used a similar procedure where members from opposing teams questioned each other to gain understanding of both positions. However, this came after each team formulated a strategy, and biases were found in comprehension of the two positions, even when members admitted to understanding both solutions.

Group versus no group. Some subjects debated their position as employers or employees in a simulated nonunionized company, while others were members of laboratory teams and representatives of abstract labor or management organizations. Since three bargaining pairs were run in one group session, there were 20 group runs and 60 individual runs. Each participant in the individual conditions was explicitly told that he was only bargaining for himself, and any decision on a settlement applied only to that particular employee and had no direct relevance for the contracts of other employees. The instructions, including the nature of the hypothetical company and the task, were essentially the same for both conditions except for modifications in wording which took into account the change in structure when individuals rather than teams competed. In fact, just two changes in format were necessitated in order that the task be made more appropriate to employer-employee interaction. Details on these procedures are discussed in the section on the task description.

Personality. Since two subject variables were used in selection, care had to be taken to avoid confounding them. Due to an apparent ideological bias in the F scale it might have been rather difficult to find high-F prolaborites or low-F antilaborites. Rokeach (1960) successfully developed a scale which is somewhat impervious to ideological content. Correlations between his Dogmatism Scale and scales of political-economic conservatism are low and nonsignificant. The D Scale was further modified by the investigator in order to remove vestiges of apparent ideological content which might have been associated with prolabor or promanagement attitudes. An attempt was also made to choose or construct items which sampled certain conceptual components of the dogmatism variable such as resisting pressures to change one's opinion, rejection of disbelievers, etc. There are 48 items fairly evenly divided over these various categories. Sixteen were borrowed from Rokeach (1960). In order to obviate an acquiescent response set from influencing the results, items are worded in two directions, reflecting both a high- and a low-dogmatic ideology. The split-half reliability, corrected by the Spearman-Brown formula, is .72.

---

[4] The coefficient of variability $\left(V = \frac{s}{\bar{x}}\right)$ rather than the standard deviation was used as a measure of dispersion for both scales. Blalock (1960, p. 73) claimed that when the means are large and spread out it might be somewhat misleading to compare the absolute magnitudes of the standard deviations.

[5] Apart from the creation of unilateral and bilateral groups, the nature of the instructions concerning appropriate behavior during these sessions may have influenced negotiating behavior. Experiments are in progress in an attempt to clarify the effect of varying the instructions for the prebargaining session both for unilateral and bilateral arrangements.

Typical of the scale are the following original items, worded in both directions:

High dogmatic	Low dogmatic
I feel quite justified in sticking to a position that I feel strongly about even in the face of strong opposition.	If progress is ever to be made in this world, we must encourage cooperation between conflicting political and religious groups.
One must be as critical as possible of the ideas of one's opponents.	A leader should look to his opponents for good ideas as well as to his supporters.

Approximately the upper and lower quartiles of the modified D-Scale distribution were eligible for participation. However, only those also scoring in the upper and lower quartiles of the labor-management scale were used. That correlated subject variables were avoided is attested to by the interscale correlation of .011. There was no problem in recruiting procompany and prounion subjects at both extremes of the modified D Scale. The scores ranged from 87 to 177, with a mean of 128.30. The mean score for high dogmatics was 143.70 ($V = 10.91$), while the mean for low dogmatics was 112.92 ($V = 10.79$).

Recruitment of Subjects

Items from the labor-management scale and the modified Dogmatism Scale were administered together. The position of a given item was randomly determined. Respondents were to indicate the extent of their agreement on a 5-point scale ranging from strongly agree to strongly disagree.[6] In order to attain enough subjects at the extremes of both scale distributions, the questionnaire was administered to 850 students from five undergraduate psychology classes at Northwestern University. Respondents were informed that we were interested in recruiting people who had similar attitudes for participation in an interesting experiment.

As soon as both scales were scored, eligible respondents were placed in one of eight categories arranged by sex, dogmatism, and prounion or procompany. Conditions were assigned to sessions each week before subjects were contacted. Both levels of the three variables were interspersed in a random-like manner in sessions during a given week.

Procedure

The situation facing the subjects was constructed in such a way so as to incorporate a number of parameters of a particular real world conflict. A complex simulation model was chosen over simpler, more popular situations on the presumption that subject behavior might come closer to their behavior in other conflict of interest situations, and the results might have more predictive validity for the situation being simulated. The simulation technique used also lends itself to the study of the effects of group commitment. A non-zero-sum collective bargaining game devised by Campbell (1960) was considerably modified for this experiment.

The following task description is for those subjects who were assigned to be representatives of laboratory teams which simulated an organization. Slight modifications, discussed below, were introduced when individuals bargained for themselves. Upon entering the laboratory the participants were told that they were being given the opportunity to learn about the bargaining process through participation and informed of research evidence indicating that simulations of processes were a more effective learning device than attending lectures or reading case studies.

Participants were then assigned to a three-member union or company team. They were told that assignment was made specifically on the basis of attitudes toward labor unions and management, but similarity on other more general attitudes was also used as a criterion. The hypothetical company was called the Acme Steel Corporation. They were told that negotiations had been going on for awhile, but no contract had been signed. They were the newly chosen representatives. Both teams received four issues, each arranged on a 17-step scale with the position of the union and company, when negotiations broke off, at either extreme. Between these extremes some possible compromises were listed for the convenience of the bargainers. Below each position on the scale the estimated cost to the company in thousands of dollars for the duration of the contract, if the issue were settled at that position, was listed. In the individual sessions the cost was in dollars, but the scale was identical (e.g. $320 rather than $32,000). An example of an issue scale is as follows:

---

[6] Both scales have been deposited in their entirety with the American Documentation Institute. Order Document No. 9408 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C. 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

	Percentage of company payment[7]									
Company	0 (0)	10 (4)	20 (8)	30 (12)	40 (16)	50 (20)	60 (24)	70 (28)	80 (32)	Union

An issue was chosen on the basis of relevance to broader problems or a widely publicized grievance and whether it could be quantitatively defined. The scale range for each issue was as follows:

Issue	Company position	Union position
Wages	No raise	$ .16 raise per hour
Off-Job training	No vocational school payment	80% tuition fees paid by company
Hospital plan	20% premium payment	100% premium payment by company
Paid vacation	2 weeks for 1 year of service (no increase for more years)	3 weeks for 10 years of service (scale started at 3 weeks for 25 years)

In terms of cost to the company and union demands, the wage issue was the most expensive, while paid vacation was the least expensive. The other two were equally expensive and fell in between.

In order to give the bargainers some perspective and to provide a common starting point for position formation, a page of background information was provided. A brief statement of company and union rationale for their demands and "going rates" at four other similar hypothetical companies in the industry were included. The average going rate on each issue was between the starting union and company positions at Acme. Such information as profit and loss statements, company budgets, union funds, and a detailed history of company operations was not provided in an attempt to insure that the results would not be a function of built-in parameters and so that the teams would be given sufficient latitude to be flexible or rigid.

The experiment was divided into two phases. The first phase consisted of either studying the issues on a bilateral basis or planning strategies unilaterally. Participants in both conditions were not told of the exact procedure to be used in the bargaining session. Before the next phase all participants were asked to privately rank the importance of the four issues. The second phase consisted of randomly dividing the six bargainers into three union-company pairs for contract negotiations. In order to reduce the potential influence on any dyad coming from another pair or to insure independence, the pairs were separated into three corners of the room, and screens were placed around each bargaining table. To make sure that team identification was maintained, the representatives were told that the results of their negotiations would be included with the results of the bargaining efforts of their partners, and net gains in terms of a total score for the team would be computed. Thus each subject was bargaining for his team and, perhaps, expected to be measured against how well his two colleagues fared. This competitive set was pitted against the advantages of a quick settlement. The latter pressure was created by informing the bargainers that each 5 minutes of debate time would cost $5,000 (or $50 in the no-group condition) to each side in lost wages or profits.

The intergroup phase lasted for 30 minutes and was divided into 5-minute intervals, each 5 minutes representing a simulated day. Announcement of each 5 minutes of debate time was made by the experimenter. The latter's role was that of "referee," but he did not specifically observe each bargaining pair. In order for a contract to be settled, agreement must have been reached on each of the four issues. Bargainers were provided with contracts. If an issue was settled, they were asked to circle the position of settlement or write it in if it fell between two listed positions and initial their contract opposite the scale. If an issue was not settled at the deadline, the bargainers were asked to indicate, independently, the amount of movement that they were willing to concede in order to obtain an agreement.

Bargaining pairs were instructed to provide some indication of the approximate amount of time needed to discuss each issue. Each bargainer was given a time sheet arranged such that he would only have to record the time on his watch at the start of discussion of an issue and then again when ready to move on to another issue. If two or more issues were being discussed simultaneously, they were to record the same time for each of the issues. Many dyads did discuss issues simultaneously, and "package agreements" were occasionally negotiated. Bargainers did not have to settle an issue to move on to another one. They could go back to unsettled issues.

At the end of the second phase all representatives were brought together to fill out postnegotiation-reaction questionnaires. Some of the questions included the defensibility of one's starting position on each issue, the degree to which one wanted to see his team (or himself) come out favorably, and the extent to which compromise was perceived as defeat. None of the subjects was sophisticated enough about the manipulations to have his data invalidated.

---

[7] This is a 17-step scale insofar as bargainers were able to use intervals of 5%.

While the participants were filling out the post-negotiation questionnaire, the experimenter combined the results for all dyads in order to give each team an idea of how it fared relative to the other. The resulting figures indicated how far the company and union teams were willing to move from their starting positions. The computational procedure is discussed in Druckman (1966).

RESULTS

It was assumed that the firmness of the two negotiators' commitment to their respective positions would be reflected in four measures of conflict resolution: the speed of resolution, average distance apart, average amount of yielding, and number of unresolved issues. Since the behavior of one representative in a bargaining dyad was not independent of the behavior of his opponent, the unit of analysis was the bargaining dyad rather than the individual bargainer. A special attempt was made to insure the independence of the three dyads in a group session during bargaining. Any given dyad was kept virtually unaware of the progress, situation, or completion of any other pair of bargainers. Since the independence of issues during debate is questionable, the effect of the manipulated conditions was assessed across the four issues.

Due to the imposition of a deadline on bargaining, there were both settled and unsettled contracts. Since it is hypothesized that the outcome of bargaining is a function of the manipulated conditions, both settled and unsettled contracts must be given equal consideration in any statistical analysis. The technique for inclusion of all contracts is specific to the measure used and is discussed below.

Speed of Resolution

A reciprocal transformation was performed on the time measure in order to insure stability in the variances (Edwards, 1960) and so that deadlocks could be taken into consideration. Deadlocks were assumed to be settled in an infinite amount of time. The reciprocal of infinity is zero so that deadlocks could be treated as zero values and included in the distribution with the reciprocals of all the other obtained values. Thus all unsettled contracts received the same score.

The harmonic mean times (the reciprocal of the mean of the reciprocals of the original values) are presented in Table 1. Bass (1963) claimed that the harmonic mean time can be interpreted as the speed with which the conflict is resolved, while the mean time is the length of time required to resolve the conflict. The harmonic means are somewhat inflated due to deadlocks. Table 2 reports the results of an analysis of variance performed on the reciprocals of the time measure. The strategy-study effect is highly significant. While the mean difference between high and low dogmatics is large and in the predicted direction, the effect does not quite reach significance. The group-no group main effect and the interactions account for a negligible portion of the total variance.

Average Distance Apart and Average Yielding

The indication of how far a bargainer was willing to go in order to obtain agreement on unsettled issues made it feasible to include all contracts, resolved and unresolved, in determining the average distance apart across all issues as well as the average yielding per contract.[8] Since each scale had the same number of steps, it was possible to equate yielding from one scale to another.

Since the results for these two measures were highly similar, they are reported together. The results of analyses of variance for both measures, as reported in Table 2, show two main effects to be significant. High dogmatics were significantly further apart in their positions at the end of the bargaining session and yielded significantly less than low dogmatics. Strategy experience kept bargainers significantly further apart and resulted in significantly less yielding than study experience. While bargainers who represented groups were further apart and yielded less than bargainers who represented themselves, the mean difference was not large enough to be significant. The means are reported in

---

[8] Yielding to compromise and the amount of movement that one is willing to concede in order to obtain an agreement were equated in obtaining the measure of average yielding. Both are indicative of a "desire to compromise" rather than "the most desirable position for settlement."

TABLE 1

MEANS FOR THREE MEASURES OF CONFLICT RESOLUTION

	Group	Strategy	No group		Group	Study	No group	High dogmatic	Low dogmatic
Speed of resolution (in min.)									
High dogmatic	67.35		58.24		28.91		23.76	44.57^{a}	
Low dogmatic	32.51		37.42		23.92		23.33		29.30^{a}
Strategy		48.90^{b}		Study		24.98^{b}			
Group	38.17^{c}			No Group			35.69^{c}		
Average distance apart									
High dogmatic	2.450		1.683		1.500		.667	1.575^{a}	
Low dogmatic	.733		1.350		.300		.250		.658^{a}
Strategy		1.554^{b}		Study		.679^{b}			
Group	1.246^{c}			No Group			.988^{c}		
Average yielding									
High dogmatic	6.942		7.208		7.250		7.667	7.267^{a}	
Low dogmatic	7.642		7.325		7.850		7.875		7.673^{a}
Strategy		7.279^{b}		Study		7.660^{b}			
Group	7.421^{c}			No Group			7.519^{c}		

[a] Means for dogmatism main effect.
[b] Means for strategy-study main effect.
[c] Means for group-no group main effect.

Table 1. None of the interactions for either measure approached significance. When issues were included as a classification category, for both measures, neither the issue main effect nor any Issue × Condition interaction was significant. Thus high scores on the modified Dogmatism Scale and strategy prenegotiation experience kept bargainers further apart and produced less yielding irrespective of the issue being bargained for.

Number of Unresolved Issues

Each contract was represented by a score ranging from 0 to 4, depending upon how many issues were not settled. The Mann-Whitney *U* test, with a correction for ties, was used to analyze this data for each variable. A feature of this test is that the effect of tied ranks on p is practically negligible.

Table 3 shows the number of bargaining dyads or contracts with 0, 1, 2, 3, or 4 issues unresolved for each condition. Strategy experience before bargaining led to a higher rate of unresolved issues per contract than bilateral study experience. A Mann-Whitney *U* test yielded a z of 2.53 which is highly significant ($p < .005$). High dogmatics resolved fewer issues, on the average, per contract than low dogmatics. A z of 2.40 is highly significant ($p < .008$). Whether an individual represented himself or a group did not seem to have an effect on how many issues were resolved. A z of .608 is not significant.

TABLE 2

ANALYSES OF VARIANCE OF THREE MEASURES OF CONFLICT RESOLUTION

Source	Reciprocal of time to complete contracts		Average distance apart		Average yielding	
	MS	*F*	*MS*	*F*	*MS*	*F*
Dogmatism (A)	.00211	2.30	25.208	5.34*	4.951	4.46*
Strategy-Study (B)	.00965	10.54**	22.969	4.86*	4.361	3.93*
Group-No Group (C)	.00009	<1	2.002	<1	.288	<1
A × B	.00058	<1	.352	<1	.000	<1
A × C	.00030	<1	8.802	1.86	1.782	1.60
B × C	.00019	<1	1.008	<1	.453	<1
A × B × C	.00000	<1	.675	<1	.069	<1
Within replicates	.00092		4.722		1.111	

* $p < .05$, $df = 1/112$.
** $p < .01$, $df = 1/112$.

TABLE 3

NUMBER OF BARGAINING DYADS WITH 0, 1, 2, 3 OR 4 UNRESOLVED ISSUES

	No. unresolved issues				
	0	1	2	3	4
Strategy	31	6	9	4	10
Study	43	7	4	3	3
High dogmatic	31	7	8	5	9
Low dogmatic	43	6	5	2	4
Group	38	8	5	4	5
No group	36	5	8	3	8

Role Differences in Yielding Behavior

Dyadic analyses could not take into account differential bargaining behavior as a function of assigned role or the defensibility of the builtin positions. Thus, a five-way analysis of variance was performed with the intention of examining differential yielding as a function of assigned role and issue rather than in the effect of the manipulated conditions per se. Role and the Role × Issues interaction were highly significant and together accounted for 78% of the controlled variance. The Dogmatism × Role interaction was not significant indicating that high or low dogmatics behaved the same irrespective of their assigned role.

The mean amount of yielding for each role by issue across the three manipulated conditions is presented in Table 4. Observation of the means reveals that the role main effect seems to be a function of differential yielding on one issue—hospital plan. The difference between the average yielding by company or employer representatives and union or employee representatives across the four issues is 2.329. When the hospital plan issue is excluded, the mean difference drops to .197.

Reference to data on perceived defensibility of the built-in starting positions by company or employer and union or employee representatives may provide an explanation for the Role × Issue interaction. Participants were to place each issue on a 5-point scale ranging from "more defensible than other side" to "not as defensible as other side." The position endorsed by the union representative or employee was subtracted from the position endorsed by the company representative or employer in each dyad. The higher the mean across the 120 contracts in a positive direction, the more defensible was the union's position perceived to be; the higher the mean in a negative direction, the more defensible was the company's position perceived to be. The mean for each issue is as follows:

Wages	Off-Job training	Paid vacation	Hospital plan
+.4298	+.1974	−.2818	−.6360

The difference between the means is highly significant ($p < .001$). Application of Duncan's multiple-range test indicated that only the mean difference between wages and off-the-job training is not significant or near significant ($p > .10$). The hospital plan issue was thus significantly less defensible to union and employee representatives than the other three issues. It was shown that differential yielding on hospital plan was considerably greater than on any of the other issues.

Table 4 also shows that company or employer representatives yielded less on paid vacation, while union or employee bargainers yielded less on wages and off-the-job training. The union position was considered to be more defensible on wages and off-the-job training, while less defensible on paid vacation and hospital plan. Thus, in a general way, differential yielding by role corresponds to the perceived defensibility of the built-in positions.

Data were not collected concerning reasons for defensibility ratings. However, a general correspondence between the average going rates and the defensibility ratings may be noted. The average going rate for wages and off-the-job training was half the distance between the company and union starting positions. The average going rate for the paid-vacation issue was four steps on the scale from the company starting position, while for hospital plan it was only three steps from the company position. The presumed relation between defensibility, going rates, and differential yielding is more suggestive than con-

TABLE 4

MEAN YIELDING BY COMPANY OR EMPLOYER AND UNION OR EMPLOYEE REPRESENTATIVES BY ISSUE

	Wages	Off-Job training	Paid vacation	Hospital plan	Mean by role
Company or employer	7.750	7.742	6.675	3.058	6.306
Union or employee	7.475	6.850	8.433	11.783	8.635
Mean by issue	7.613	7.296	7.554	7.421	

clusive. As such these data suggest hypotheses which may be tested in another setting by, perhaps, manipulating defensibility and determining amount of yielding.

Agreement on the Ranked Importance of the Issues

If the study session was effective in facilitating communication between bargainers, this should be reflected in agreement on the ranked importance of the four issues. The method for determining the extent of agreement between two members of a bargaining dyad was to take the absolute difference between the ranks assigned to each issue by each member and sum the differences. The "agreement scores" were ranked and a Mann-Whitney U test was performed. The hypothesis was confirmed. Bargainers who studied the issues before negotiation agreed more on their relative importance than bargainers who planned strategies apart from one another. A z of 2.23 is highly significant ($p < .01$). The difference in amount of agreement between high dogmatics and low dogmatics or between group and individual representatives was slight and nonsignificant.

The amount of time spent discussing each issue corresponded to its average rank in importance. The wage issue was considered to be the most important ($\bar{x} = 1.26$) and was discussed longest in debate (8.15 minutes), off-the-job training was second in both importance (2.19) and time to discuss (7.18 minutes), followed by hospital plan (3.03 and 4.13 minutes) and paid vacation (3.52 and 3.65 minutes). The order of ranked importance was not affected by condition or role.

Involvement and the Perception of Compromise as Defeat

In general, representatives wanted themselves or their teams to come out favorably in the negotiations. The overall mean was 1.45 on a 4-point scale. Group representatives wanted their team to come out favorably significantly more than individuals wanted themselves to come out favorably. The means were 1.37 in the case of group representatives and 1.54 for bargainers representing themselves ($t = 2.51$, $p < .01$). The differences between the other conditions were small and nonsignificant.

High dogmatics viewed compromise as defeat significantly more than did low dogmatics. The question was asked on a 3-point scale. The means were 2.11 for high dogmatics and 2.37 for low dogmatics ($t = 3.36$, $p < .001$). The differences between the other conditions were small and nonsignificant.

DISCUSSION

The results appear to be quite consistent across the measures of conflict resolution used. Strategy experience before bargaining led to a hardening of positions as reflected in measures of agreement and amount of yielding. Bilateral study experience, on the other hand, resulted in faster agreement and more yielding on the part of bargainers. These results were a function of prenegotiation experience apart from the effects of group commitment per se. Whether bargainers represented groups or themselves did not have a significant effect on compromise behavior. That study experience was more effective in facilitating communication may have been revealed in the greater agreement on the ranking of the importance of the issues by bargainers after bilateral study.

The modified Dogmatism Scale was also predictive of conflict resolution. Irrespective of strategy or study prenegotiation experience, high dogmatics were more resistant to compromise than low dogmatics. A postnegotiation question revealed that high dogmatics viewed compromise as defeat more than low dogmatics. The dogmatic subjects behaved in keeping with the theory and evidence presented by Rokeach (1960) and the evidence of Jamias (1964). They were less willing to defect from given positions. Blake and Mouton (1962) and Sherif and Sherif (1965) dismissed the relevance of psychodynamic explanations of behavior in an intergroup situation. The results of this study suggest that these explanations cannot be dismissed so lightly. The low correlation between the modified Dogmatism Scale and the labor-management attitude scale suggests that the former may be tapping a general disposition not related to political or specific content attitudes. Also there were no differences between high- or low-dogmatic union representatives and high- or low-dogmatic company representatives on yielding behavior. It might be interesting to examine the relationship between this scale and extant measures of misanthropy.

Bargainers who represented groups behaved no differently than those who bargained for themselves. A careful attempt was made to maintain group identification during the bargaining session. Bargainers were told, before negotiations, that a team score would be determined and a "winner" and "loser" announced. The results of a postnegotiation question indicated that this manipulation might have been effective. Group representatives wanted their team to come out favorably more than individuals wanted themselves to come out favorably. It is conceivable that such alternative procedures as having all three representatives bargain together as a team or choosing one to represent the team during bargaining might have intensified the salience of representing a laboratory team. However, both alternatives would require a greatly increased number of group sessions and the possibility of unwanted constraints on the bargainers from the bargaining session itself due to the "public" nature of the situation. Only further investigations can clarify this issue.

Experimental studies of intergroup relations have been characterized by the use of short-lived ad hoc groups. The Sherif et al. (1961) boys' camp study is perhaps the only exception. While certain steps were taken in this study to increase group cohesion, group identification or the intensity of competition might have been increased if the groups interacted over a longer period of time. At the current stage of empirical development it is difficult to assess the relevance of findings from simulated groups for groups with their own developmental history.

The implications of this study are clear. While the negotiation situation facing professional diplomats and bargainers is more complicated and more is at stake, these results may be suggestive to those social engineers responsible for creating situations more conducive to agreement. The unilateral attachments of representatives may not lie in their group identifications, but may be the result of a disciplined reticence toward their opponents. When representatives are given positions to defend but compromise is the desired goal, a problem-solving atmosphere, with no sanctions against open discussion from either side, is recommended. However, this should take place before representatives have a chance to develop rationale or a cognitive context for their own position. This may be similar to the situation in labor-management contract negotiations where arbitrators assign starting positions to the contending parties and have them discuss their relative merits in the context of both vantage points. It also appears that efforts aimed at reducing conflict have a better chance of succeeding if carried out by people who are more collaborative in interpersonal relations and who do not view compromise as defeat.

REFERENCES

BASS, B. M. Effects on negotiators of their prior experience in strategy or study groups. Technical Report No. 1, 1963, Contract Nonr 624(14), Office of Naval Research.

BASS, B. M., & DUNTEMAN, G. Biases in the evaluation of one's own group, its allies and opponents. *Journal of Conflict Resolution*, 1963, 7, 16–20.

BLAKE, R. R. Psychology and the crisis of statesmanship. *American Psychologist*, 1959, 14, 87–94.

BLAKE, R. R., & MOUTON, J. S. Competition, communication, and conformity. In I. A. Berg & B. M. Bass (Eds.), *Conformity and deviation*. New York: Harper, 1961. Pp. 199–229.

BLAKE, R. R., & MOUTON, J. S. The intergroup dynamics of win-lose conflict and problem-solving collaboration in union-management relations. In

M. Sherif (Ed.), *Intergroup relations and leadership*. New York: Wiley, 1962. Pp. 94–140.

BLALOCK, H. M. *Social statistics*. New York: McGraw-Hill, 1960.

CAMPBELL, R. J. *Originality in group productivity: III. Partisan commitment and productive independence in a collective bargaining situation*. Columbus: Ohio State University Research Foundation, 1960.

CROWNE, D. P. Family orientation, level of aspiration, and interpersonal bargaining. *Journal of Personality and Social Psychology*, 1966, **3**, 641–645.

DEUTSCH, M. Trust, trustworthiness, and the F scale. *Journal of Abnormal and Social Psychology*, 1960, **61**, 138–140.

DEUTSCH, M. Cooperation and trust: Some theoretical notes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1962*. Lincoln: University of Nebraska Press, 1962. Pp. 275–320.

DRUCKMAN, D. Dogmatism, prenegotiation experience and simulated group representation as determinants of dyadic behavior in a bargaining situation. Unpublished doctoral dissertation, Northwestern University, 1966.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Holt, Rinehart & Winston, 1960.

FERGUSON, C. K., & KELLEY, H. H. Significant factors in overevaluation of own-group's product. *Journal of Abnormal and Social Psychology*, 1964, **69**, 223–228.

GLADSTONE, A. I. Relationship orientation and processes leading toward war. *Background*, 1962, **6**, 13–25.

JAMIAS, J. F. The effects of belief system styles on the communication and adoption of farm practices. Unpublished doctoral dissertation, Michigan State University, 1964.

LUTZKER, D. R. Internationalism as a predictor of cooperative behavior. *Journal of Conflict Resolution*, 1960, **4**, 426–430.

MACK, R. W., & SNYDER, R. C. The analysis of social conflict—toward an overview and synthesis. *Journal of Conflict Resolution*, 1957, **1**, 212–248.

PEARSON, L. B. *Diplomacy in the nuclear age*. Cambridge: Harvard University Press, 1959.

ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.

RUSK, D. Parliamentary diplomacy—debate versus negotiation. *World Affairs Interpreter*, 1955, **26**, 121–138.

SHERIF, M., HARVEY, O. J., WHITE, B. J., HOOD, W. R., & SHERIF, C. W. *Intergroup conflict and cooperation: The robbers cave experiment*. Nórman, Okla.: University Book Exchange, 1961.

SHERIF, M., & SHERIF, C. W. Research on intergroup relations. In O. Klineberg & R. Christie (Eds.), *Perspectives in social psychology*. New York: Holt, Rinehart & Winston, 1965. Pp. 153–177.

SUMNER, W. G. *Folkways*. Boston: Ginn, 1906.

TURK, H., & LEFCOWITZ, M. J. Towards a theory of representation between groups. *Social Forces*, 1962, **40**, 337–341.

VEGAS, O. V., FRYE, R. L., & CASSENS, F. P. Learning set as a determinant of perceived cooperation and competition. *American Psychologist*, 1964, **19**, 482. (Abstract)

(Received July 20, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 291-303

PERSONALITY DETERMINANTS OF PSYCHOPHYSIOLOGICAL RESPONSE TO STRESS:

A THEORETICAL ANALYSIS AND AN EXPERIMENT[1]

EDWARD M. OPTON, JR. AND RICHARD S. LAZARUS

University of California, Berkeley

Personality correlates of psychological and physiological response to 2 different laboratory stressors were sought. Item analysis of a 795-item personality inventory revealed 131 discriminating items in an ipsative (intraindividual) design. Ss who responded relatively more strongly to a threatening motion-picture film than to threat of electric shock described themselves as lacking in impulse expression, socially inhibited, introverted, submissive, suggestible or obedient, insecure, passive, anxious, and not caring about friends. Ss who responded more strongly to threat of shock than to the movie attributed the opposite qualities to themselves. Normative (interindividual) analyses (high response to film compared to low response to film, high response to shock compared to low response to shock) found no more discriminating personality items than would be expected by chance. Both the results of this experiment and a theoretical analysis of the problem indicate that ipsative research designs can control much of the extraneous variance which probably is responsible for the weak and ephemeral relationships between personality and psychophysiological response commonly found in studies using normative designs.

One who seeks personality determinants of autonomic response assumes that the peripheral response mirrors the central psychological event; that, for example, a vigorous vasoconstriction in the finger indicates a more important psychological event than a mild vasoconstriction. The assumption may often be valid if we compare two different events in a single person, events whose significance is measured at the periphery of a single individual's nervous system. But if the comparison is between individuals, then individual differences in neuroanatomy or neurophysiological functioning will be very important sources of error. These sources of psychophysiological variation not relevant to personality tend to be overlooked by psychologists, who sometimes forget that "the" autonomic nervous system varies quite widely among individuals (Jost & Sontag, 1944; Pavlov, 1927, Ch. 17).

Unfortunately, there is no way to assay separately the psychologically irrelevant constitutional differences among individuals' nervous systems. The usual statistical partitioning of psychophysiological response variance into general activation, individual response specificity, stimulus specificity, and correlation with prestimulus level does not help, for each of these statistically derived portions of variance represents both psychological and constitutional effects. Hence, if one is working with between-subjects (normative) comparisons, the psychologically irrelevant constitutional differences have to be treated as error variance. That this error variance is almost certainly very considerable is suggested by the following facts: (*a*) Large differences in autonomic response exist from birth (Grossman & Greenberg, 1957; Richmond & Lustman, 1955), and individual differences in neonates could hardly be psychologically determined; (*b*) large differences in the end organs are known to exist in some autonomic pathways—for instance, the number of active sweat glands on the palm varies greatly among individuals (Darrow, 1927, p. 213); (*c*) psychophysiological levels and responsivity vary considerably among individuals even during sleep (Craig, 1936; Titelbaum,

---

[1] This study was supported by Research Grant MH-2136 from the National Institute of Mental Health of the United States Public Health Service to the authors and by Training Grant No. RH4 from the Vocational Rehabilitation Administration, Department of Health, Education, and Welfare. The study is based on data collected and analyzed by Elizabeth Alfert in her doctoral dissertation (Alfert, 1966).

1941); (*d*) there is a very large hereditary component in a popular autonomic measure—skin-conductance reactivity (Lehtovaara, Sarrinen, & Järvinen, 1965).

In spite of the problem of excessive error variance, there has been no lack of effort to find physiological correlates of psychological traits.[2] For the most part, these efforts have yielded disappointingly slight and ephermeral relationships between personality and psychophysiological reaction.

This article stems from efforts to find connections between personality and individual differences in response to stress. We shall first offer a logical analysis of the problem and a methodological explanation for the dearth of stable findings; then we shall describe an experiment in which the first attempt to obtain a relationship failed, while a second succeeded.

WHY SHOULD PERSONALITY BE RELATED TO STRESS RESPONSE?

Imagine an experiment in which the subjects are exposed to the threat of a painful electric shock, and in which skin conductance and heart rate are measured before the threat is induced and during the period while the subjects are awaiting the shock. As everyone who has worked with such situations and variables knows, there will be large differences among subjects in the amount of autonomic disturbance that is observed. Some will show very marked reactions, others minimal or perhaps even no reaction. It is commonly assumed that these variations are not random, and that differences in the subjects' personalities account for some of the variation. But what kinds of personality variables are actually relevant? Why should personality be associated with the response variation? This is the key theoretical question which must be addressed.

Actually, there are many ways in which personality variables could affect response to the threat of electric shock. It would not be germane to our purpose to review all the possibilities, but a few examples will help clarify the problem. The subjects' attitudes toward pain may be involved. To some, pain may be more feared, and so the threat may be more threatening. In taking a stoic or a fearful attitude towards the prospect of pain, subjects may differ, yet maintain conformity to the norms of their respective cultures or subcultures (Sternbach & Tursky, 1965; Zborowski, 1952). Other subjects who show a large reaction may not be reacting to the fear of pain so much as to social or symbolic aspects of the situation—to some it may symbolize a punitive father; to some it may threaten to expose personal inadequacies; to some manipulation by the experimenter may produce an unpleasant feeling of helplessness; and to some the ambiguity of the situation and of what would be appropriate behavior for their role may constitute the major threat. Moreover, subjects may react differently to threats depending on their desire and ability to express, control, or suppress their emotions.

If situation and subjects can be controlled and matched sufficiently to predict that only one of the above factors will be operative, then the researcher need only measure the relevant personality characteristic and correlate it with observed stress reaction. But if multiple personality factors are relevant, the problem becomes extremely complex and difficult to study. The basic methodological difficulty is the presence of unwanted or irrelevant sources of variance in the measured stress response—irrelevant, that is, to the particular personality trait being studied.

IRRELEVANT SOURCES OF VARIATION IN THE STRESS RESPONSE

If a subject is exposed to a particular stress condition, the magnitude of his autonomic response is, in part, a function of the prestress level. This problem, sometimes called

---

[2] The between-subjects (normative) experimental design has been used by psychophysiologists in conjunction with a variety of trait measures, for example, MMPI-derived scales of ego strength (Roessler, Alexander, & Greenfield, 1963), adjective check lists, *Q* sorts, and the California Psychological Inventory (Block, 1957), personality ratings (Seymour, 1950), Rorschach indexes of emotionality (Lacey, Bateman, & Van Lehn, 1952), interview-based estimates of various personality attributes (Darling, 1940), interview-based ratings of affective lability (Oken, Grinker, Heath, Herz, Korchin, Sabshin, & Schwartz, 1962), scores on achievement and ability tests (Lacey, Kagan, Lacey, & Moss, 1963), observer ratings (Jones, 1950), and questionnaire measures of introversion-extraversion (Eysenck, 1956; Lynn & Eysenck, 1963).

"the law of initial values," is well known, and is usually dealt with by making a covariance adjustment based on the correlation of the prestimulus levels with the stress levels (Benjamin, 1963; Lacey, 1956). Even after this adjustment, however, another problem remains. It is still not clear to what extent interindividual variations in autonomic response are due to the stimulus properties of the particular stress condition or to psychologically relevant dispositions on the part of the individual to give strong reactions to this type of stress condition, or to constitutional properties of the individual that have little to do with psychological processes. For example, the individual with a highly reactive autonomic system or, more specifically, a labile heart response, might show high levels of activation in heart rate, not because he is particularly disturbed psychologically by that stress stimulus, but because constitutionally he responds to any stimulus with marked activation. In short, a major source of variation that may have nothing to do with personality accounts for his response. If this source of variation is substantial, it will tend to obscure any correlation with personality variables.

We said, ". . . a major source of variation that *may* have nothing to do with personality . . .," because it is possible that a consistent pattern of high reactivity could result from personality factors just as easily as from constitutional properties of the person. The person might react strongly not because of some physical property of his constitution, but because of great psychological sensitivity to many situations. Unless the constitutional or psychological property assumed to be operating is independently identified, one cannot assert which of these two classes of variables accounts for the stable response tendency. Still, the main point is that constitutional characteristics probably play a substantial role in level of autonomic response. To the extent this is so, the relationship of stress response to personality will be obscured by an irrelevant source of variation.

There are two additional irrelevant sources of variation, stimulus specificity and individual response specificity, that will also tend to obscure personality–stress-response correlations. Increasing attention has been directed at these concepts, and they have become well established in empirical research in psychophysiology (e.g., Engel, 1960; Lacey, 1959; Lacey et al., 1963).

Stimulus specificity is the tendency for certain classes of arousing stimuli to result in particular forms of autonomic reaction. There is evidence, for example, that one type of stimulus condition will tend to produce heart-rate acceleration while another will produce deceleration (Lacey et al., 1963). Stimulus specificity thus identifies a source of variation (the stimulus) which obscures the correlation that might exist between personality variables and degree of autonomic response to various stress conditions.

Individual response specificity is the tendency of subjects to display individual, stable patterns of autonomic reaction across a variety of stress conditions. For example, one individual will tend to react primarily with rises in heart rate or blood pressure, while another will show activation in skin conductance or respiration. Again we have an example of a source of variation in the response which may be unrelated to personality. To the extent that such sources of variation are independent of personality, the correlation between personality and stress response will be more difficult to demonstrate because the irrelevant, constitutional source of variation pushes the specific autonomic response measure up in one case and not in another.

Here too, as noted before, one must not take empirical evidence of individual response specificity as necessarily signifying constitutional sources of variation. The stable pattern of reaction could just as easily be the result of personality dispositions to interpret certain classes of stimuli in a particular way. Stereotyped patterns of autonomic reaction could result from psychological intervening variables rather than from the physical structure of the autonomic nervous system or its end organs. In such a case, if the investigator made his hypothesis in terms of theoretically appropriate personality variables, response specificity would not be an irrelevant source of variation. However, in the usual case, when the investigator's hypothesis about personality is not predicated on a response-specificity hypothesis, then individual response specificity will be an irrelevant source of variation even if it is not altogether constitutional in nature. The fact that a different and unknown personality trait may be involved does not help the investigator to track down the personality correlates of stress response if his hypothesis and resulting choice of variables are based on an entirely different concept, such as the specific threat value of a situation or the ego-control processes as regulators of amount of activation. It is simply too probable that constitutional variables are involved in individual response specificity to ignore this personality-irrelevant source of variation in the stress response.[3]

HOW TO RULE OUT THE PERSONALITY-IRRELEVANT SOURCES OF VARIATION

Since major irrelevant sources of response variation operate in experiments, it is no wonder that such relationships with personality as have been found have been slight and evanescent. The irrelevant variables must be either eliminated or controlled to permit the personality correlates, if they exist, to stand out. This can be accomplished if the research design is partly intraindividual or *ipsative*, rather than purely interindividual or *normative*. Let us consider this point more closely.

In contrast with the procedure of exposing two or more groups of subjects to a single stress condition each, an ipsative or intraindividual design calls for exposing all subjects to more than one stress condition, and for determining the extent to which their reactions to these conditions are selective. The question is not whether a given subject with some trait of personality is highly or minimally aroused under some one stress stimulus, but whether he reacts more strongly to one than to another. This ipsative approach was used by Vogel, Raymond, and Lazarus (1959) in a study of the role of motivation in making particular situations stressful, and by Lacey et al. (1963) to demonstrate individual response specificity. Lacey et al. exposed each individual to several different stress conditions in order to study the individual-consistent patterns of response across these situations, that is, the extent to which a subject shows the same pattern in two or more stress conditions. The same design can be used to seek differences in reactions to different stress conditions. If we are interested in personality correlates, we ought to ask whether a particular kind of person is more disturbed by (or reacts differently to) one stress condition than another.

Such an intraindividual perspective makes it easier to speculate about which personality variables might be involved in stress response. But equally important, it has a strong methodological argument in its favor. Indiscriminate arousal may have personality determinants, but it is also confounded with all the personality-irrelevant sources of response variation we spoke of above. Selective arousal is much less confounded, since the individual serves as his own control, and constitutional and other irrelevant dispositions are held constant over several stress conditions. Selective reaction is, therefore, the prime context for the discovery of the personality correlates of stress response.

One final point needs to be made. We have cast the discussion of normative and ipsative analysis of stress response in terms of autonomic variables. However, similar arguments apply to behavioral measures such as self-reported anxiety. It is unlikely that the main irrelevant sources of reported anxiety lie in constitutional variables. However, dispositions to respond in particular ways to the obligation to give reports of one's internal state may be irrelevant to the personality correlates of situation-specific stress reactions. To the extent that the individual's report reflects, for example, yeasaying or naysaying, the tendency to use extreme or central responses, or the desire to put himself in a favorable light, say as emotional or unemotional, the correlations between verbal stress response and personality will be reduced. The ipsative design should be useful when the dependent variables are verbal or behavioral, as well as when they are physiological.

These same arguments apply with equal force to the clinical problems of psychosomatic disease. The excessive, chronic psychophysiological reactions that produce psychosomatic diseases are no doubt due partly to constitutional factors and partly to personality and situational factors. The prospects of isolating these sources are increased by research designs which hold constant or rule out one or the other of these factors. The present analysis of the advantages of the intraindividual design applies to psychophysiological research in general, rather than to any specific reaction.

---

[3] It is also quite possible that constitutional differences *are* related to personality. However, even if this were so, such constitutional differences would remain a source of error variance unless they were related to the particular personality traits under investigation.

AN ILLUSTRATIVE EXPERIMENTAL STUDY

The procedures and data to be described were incidental to another study reported by Alfert (1966) comparing subjects' stress reactions in a vicarious and direct-threat situation. The original purpose of the study was to evaluate the relationship between an individual's reaction to a stressful motion-picture film and a situation in which he was directly threatened, since if there were not consistency in reaction across two types of conditions, doubt would be thrown on the validity of the motion-picture-film approach as an analogue of life threats. Since Alfert used two different stress situations in a repeated-measurements design and also obtained data on personality, it is possible to analyze her data so as to compare the merits of the normative and ipsative methods of isolating the relationship between personality and stress response. The detailed procedures of Alfert's study have been described in the published report, so they need only be summarized here.

Forty-eight male undergraduate college students were exposed to two experimental treatments, a film called *It Didn't Have to Happen* and the threat of electric shock. The film pictured three rather harrowing accidents in a wood mill, the latter two of which could be anticipated by the viewer before they happened. The film produced considerable elevation in heart rate and skin conductance when the accidents were anticipated; response reached a peak just beyond the point of impact, then declined toward the basal level. Anxiety, measured by the Nowlis Adjective Check List of Mood (Nowlis & Nowlis, 1956), also increased substantially following the film. The shock-threat treatment consisted of telling the subject he would receive a shock following the last of a series of seven clicks at 5-second intervals. Actually, the shock was barely perceptible; the threat was in the anticipation. During the shock-anticipation period, heart rate, skin conductance, and self-reported anxiety increased for almost all subjects. The film was presented first for half of the subjects; the threat of shock came first for the other half. The results were not significantly affected by the order.

Skin conductance and heart rate were recorded continuously and measured at 10-second intervals. Stress response was calculated as the difference between the mean of the readings during the 3-minute base line preceding each stressor and the maximum value reached during the stress period.

The Nowlis check list was administered at the beginning of the experiment, after the film, and after the threat of shock had been announced.

The personality measure was a 795-item, true-false questionnaire, which the subjects filled out before entering the laboratory. The questionnaire included the 500 items of the Berkeley Psychological Inventory[4] (BPI), the *Hy*, *Pt*, *Pd*, *Hs*, *A*, and *R* scales of the MMPI, the Impulse Expression Scale (IES) of Sanford, Webster, and Freedman (1957), an adjective check list based on Mandler, Mandler, and Uviller's (1958) Autonomic Perception Questionnaire (APQ), items from Aas' (1962) Subjective Experience Inventory (SEI), and a few miscellaneous items (MIS). This questionnaire was the basis for what was frankly a fishing expedition for personality correlates of the stress reactions to the two experimental treatments.

Both normative and ipsative item analyses were carried out. In the normative analysis the two experimental treatments were dealt with separately: first, the reactions to the film were used to identify items that discriminated between high and low reactors; then reactions to the threat of shock were used in the same way. In the ipsative analysis the dependent variable was the *difference* in response to the two experimental treatments regardless of absolute levels of response. Thus, in the ipsative analysis items were identified which discriminated subjects whose response to the film was greater than their response to the shock threat from subjects who were relatively more threatened by the shock threat than by the film.

Normative Analyses

Responses to the stress stimuli were adjusted by a covariance technique (for details, see Alfert, 1966) to free them from correlation with basal levels. The corrected change scores were then used to divide the subjects at the median for each dependent variable for each stress stimulus. In other words, subjects who showed high (above median) and low (below median) heart-rate response to the stressful film were identified; similarly, groups high and low in skin conductance and in self-reported anxiety were constituted. The same procedure was used to identify high- and low-response groups for the shock threat.

Six separate item analyses were performed on the questionnaire: one analysis for heart-rate response to the film, one for heart-rate

---

[4] The authors appreciate the assistance of Jack Block, University of California, Berkeley, in making this unpublished test available to us.

TABLE 1

NUMBER OF ITEMS DISCRIMINATING CONTRAST GROUPS AT THE .05 SIGNIFICANCE LEVEL

	Criterion of response					
	Heart rate		Skin conductance		Anxiety rating	
	No. items	Significance level[b]	No. items	Significance level[b]	No. items	Significance level[b]
Normative analyses						
High vs. low response to:						
Film	10	>.50, *ns*	9	>.50, *ns*	11	>.50, *ns*
Shock threat	17	>.50, *ns*	13	.40, *ns*	16	>.50, *ns*
Ipsative analyses[a]	35	.02	8	>.50, *ns*	35	.03

[a] The contrast groups were subjects whose response to the film was greater than response to the threat of shock *versus* subject whose response to the threat of shock was greater than response to the film.

[b] Tabulated significance levels are the Monte Carlo probabilities (Block, 1960) that as many or more items would discriminate contrast groups at the nominal .05 level if the subjects were allotted to the contrast groups on a random basis.

response to the shock threat, etc. Table 1 shows the number of items found to "significantly" discriminate between high and low groups in each analysis and the probability that as many items would appear "significant" in an analysis of randomly constituted groups.

Examination of Table 1 indicates that only a few questionnaire items predicted high or low reaction to either of the two stress situations. In fact, the numbers of items producing significant discriminations were less than the number expected by chance. In view of the fact that this was a fishing expedition and that discriminating items were not predicted from any prior analysis, we can hardly regard the normative procedure as having revealed anything noteworthy. Even aside from the small number of discriminating items, we found that those items that did appear to distinguish high and low reactors did not cluster with respect to content so as to allow coherent induction about relevant personality characteristics.[5] One would at this point be forced to conclude that there are no grounds for assuming that a replication would reveal a similar pattern. Meager findings such as these abound in the research reports in this field, some a little better, some perhaps worse. (The latter are probably not reported because negative findings tend to be meaningless and do not end up in print.)

Ipsative Analysis

The ipsative procedure differed from the normative one above in that individuals' differential reactions to the two stress conditions were compared. Subjects were divided into two groups for each dependent variable (heart rate, skin conductance, and anxiety). Those subjects whose response magnitude to the film was greater than their response to the shock threat were placed in one group; those who responded more vigorously to the shock threat than to the film were placed in the other group. Thus, the division was independent of absolute response magnitude and depended only on the *relative* response to the two stress stimuli.

Assignment of subjects to ipsative groups was complicated by the existence of two measurement base lines for heart rate and skin conductance, one before the film and one before the shock threat. In 5 cases for heart rate and in 18 cases for skin conductance the classification of a subject as a "shock threat" or a "film" responder depended on whether the initial base line was used as the base for responses to both stimuli, or whether each response was measured from the immediately preceding basal value. These ambiguous cases were eliminated from the pertinent analyses.

The questionnaire responses of the remaining subjects were then item analyzed to identify the items that differentiated film responders (subjects who responded more to

[5] A complete list of discriminating items is available from the authors.

the film than to the shock threat) from shock-threat responders. The results of these analyses are summarized at the bottom in Table 1, which also shows the probability, computed by Block's Monte Carlo method,[6] that as large a number of items would have appeared "significant" by chance. Table 1 shows that the ipsative analysis produces a much larger number of discriminating items. Only in the ipsative analysis is the number of such items greater than would be expected by chance. Table 2 lists the items which discriminated at the .01, .05, or .10 significance levels.

The 131 discriminating items in Table 2 seemed to represent many different dimensions of personality. In order to determine objectively which of the possible dimensions actually discriminated the two subject groups, each author independently made a list of those personality dimensions which seemed most strongly represented among the 131 discriminating items. The dimensions were then combined into a single list of 14 bipolar personality dimensions. These dimensions head the columns in the right-hand portion of Table 2. The 131 discriminating items were presented to three psychology graduate students who were unfamiliar with the study. The order of the items was randomized so that the students would not know which items were characteristic of one or another subject group. The students were asked to assign each item to one of three positions on each bipolar personality scale. For example, a rater would decide whether a "yes" response to the item, "I find it hard to make talk when I meet new people," would be characteristic of an *anxious* person, an *untroubled* person, or *neither*, that is, not relevant. The rater would then make the same three-alternative rating about the next dimension, *insecure*, *secure*, or *neither*. The ratings of the three students were combined into a consensus by classifying an item as characteristic of one pole of a personality dimension if the three raters were unanimous in placing it there, or if two raters placed it there and the third rater placed it no farther removed than the neutral (i.e., "neither, not relevant") category.

Using the consensus ratings, 2 × 2 tables were constructed for each personality dimension. Consensus rating as characteristic of one pole or the other of the personality dimension determined placement in the left or the right column of the 2 × 2 table; endorsement significantly more often by one than by the other of the ipsative subject groups determined placement in the upper or lower row of the table. Fisher's exact test was then used to compute the probability that the ipsative subject groups differed along each of the 14 personality dimensions.

In Table 2 the personality dimensions have been listed from left to right in the order of the statistical significance of their discrimination between the ipsative subject groups. Table 2 could be summarized as follows: subjects who showed more stress rseponse to the film than to the threat of shock tended to describe themselves (in descending order of statistical significance) as: lacking in impulse expression ($p < .001$), socially inhibited ($p < .001$), introverted ($p < .001$), submissive, suggestible, or obedient ($p < .001$), insecure ($p < .001$), passive ($p = .002$), anxious ($p = .002$), not caring about friends ($p = .052$).

In addition, these film-greater-than-shock-response subjects showed nonsignificant trends to endorse items judged as (in decreasing order of approach to statistical significance): distrustful, alienated, ideational, naive, not curious, creative, or intellectual, mother centered.

The personality dimensions characteristic of the subjects who showed more stress response to the threat of shock than to the film were, of course, the opposite of those above; the dimensions may be read from the column labels of the lower half of Table 2.

---

[6] The probabilities were computed using Block's (1960) computer method of Monte Carlo determination of the expected number of "significant" items in random groups. A probability of .05 in Table 1 would mean that when subjects were assigned to groups at random, only in 5 of 100 comparisons of such random groups would the number of "significant" items equal or exceed the number found "significant" for the criterion groups.

TABLE 2

ITEMS AND PERSONALITY DIMENSIONS DISCRIMINATING IPSATIVE SUBJECT GROUPS[a]

Source		Condensed text of item	Criterion[b]	Significance level	Consensus rating of personality dimensions[c]													
Test	Item no.				Lack of impulse expression	Social inhibition	Introverted	Submissive, suggestible, obedient	Insecure	Passive	Anxious	Not caring about friends	Distrustful	Alienated	Ideational	Naive	Not curious, creative, intellectual	Mother centered
MMPI	180	Hard to talk when meet new people	HR, ANX	.10, .05	+	+	+	+	+	+	+		+	+				
BPI	226	Lower eyes when looked in face	HR, SC	.05, .10	+	+	+	+	+		+							
MMPI	265	Safer to trust nobody	HR	.10	+	+			+		+	+	+	+		−		
MMPI	351	Anxious and upset when make short trip	HR	.05	+	+			+		+	+						
BPI	397	Hard to act natural with new people	HR	.01	+	+			+		+							
BPI	450	Sentimentality embarrasses me	ANX	.10	+	+						+						
BPI	345	Do not understand those carried away by emotions	HR	.10	+		+					+					+	
IES		No man of character would ask fiance to have sexual intercourse	HR	.10	+			+						−		+		
BPI	303	Never indulged in unusual sex practices	ANX	.05	+			+									+	
BPI	86	Have dreams which disturb sleep	SC	.10	+				+		+							
APQ		When anxious or excited have upset stomach	HR	.10	+						+				−			+
BPI	294	Hard to set aside a task	HR	.05	+				+									
MMPI	358	Bad words come into mind and cannot get rid of them	HR	.05	+													
BPI	43	Stick at a job though not getting results	ANX	.10	+													
BPI	296	Had fewer close friends than most people	ANX	.10		+	+		+	+	+	+	+	+				
MMPI	305	When with people feel lonely	HR	.05		+	+		+	+	+	+		+				
MMPI	81	Would like work of forest ranger	HR	.10		+	+					+		+	−			
MMPI	530	Often afraid I am going to blush	ANX	.10		+		+	+	+	+					+		
MMPI	357	Have given up doing a thing because thought too little of my ability	ANX	.05		+		+	+	+								
BPI	417	Feel nervous and ill-at-ease at formal dance or party	HR, ANX	.01, .10		+			+		+							
BPI	289	Nervous if meet a lot of people	ANX	.10			+	+	+		+							
MMPI	171	Uncomfortable to put on stunt at party even if others doing same	ANX	.05			+			+					+			
BPI	438	Easy to drop or break with friend	HR	.05			+					+						
MMPI	141	Conduct controlled by (others') customs	SC	.10				+		+				−				
BPI	89	Like to have qualities of mother	ANX	.10				+										+
BPI	157	More pessimistic about life than most	HR	.10					+		+		+	+		−		
BPI	393	Parents should not have married	ANX	.10					+		+			+		−		
MMPI	142	Feel useless at times	SC	.01					+		+							
BPI	356	Something wrong with person who cannot take orders	HR	.10						+								

---

[a] Items characteristic (when answered "yes") of subjects responding more strongly to the film than to the threat of shock appear above double horizontal lines. Items characteristic of subjects responding more strongly to the threat of shock than to the film, below.

[b] SC, HR, and ANX (anxiety rating) were the criteria by which subjects were divided into item-analysis contrast groups. SC = skin conductance, HR = heart rate.

[c] See text for explanation of method of rating. The plus and minus signs indicate what the raters' consensus was for each item as to its relevance for each personality dimension. For instance, the + at the intersection of the first row and the first column means that the raters' consensus was that a person responding "yes" to the item, "Hard to talk when meet new people," would be likely to lack impulse expression. If the + at this intersection had been − it would have meant that the consensus of the raters was that a yes response was characteristic of the opposite pole of the dimension, free or easy impulse expression. The same analysis as is shown here was also performed by the two authors with virtually identical results. However, since the authors had prior familiarity with the items, their judgments have not been used in the statistical analysis.

Table 2—Continued

Source					Consensus rating of personality dimensions[a]													
Test	Item no.	Condensed text of item	Criterion[b]	Significance level	Lack of impulse expression	Social inhibition	Introverted	Submissive, suggestible, obedient	Insecure	Passive	Anxious	Not caring about friends	Distrustful	Alienated	Ideational	Naive	Not curious, creative, intellectual	Mother centered
MMPI	226	My family has habits which annoy me	HR	.05							+							
BPI	478	Uncertain about my beliefs and opinions	SC	.10							+							
BPI	276	Worry about responsibilities of marriage	ANX	.05							+							
APQ		Increase in muscle tension when axious or excited	HR	.05							+				−			
MMPI	93	Most people would lie to get ahead	HR	.05									+	+		−		
BPI	304	People in authority bossier than have to be	SC, ANX	.10, .05				−					+	+				
BPI	187	Life meaningless without devotion to cause	SC	.10										+	+	+		
IES		Less progress without rebellious youth	HR	.10				−						+	+		−	
BPI	432	Average person cannot appreciate art and music	HR, ANX	.05, .10										+		−		
BPI	216	When mentally engrossed, am oblivious to all else	SC	.05											+		−	
BPI	456	Life full of interesting things	HR	.10					−	−				−	+		−	
SEI	30	Have produced very odd visual experience by staring actively at something	HR, SC	.05, .05	−										+		−	
BPI	476	Strong person able to make up his mind	HR	.10												+		−
MIS		Higher rating to mother likeness	HR	.10														+
BPI	396	Argument necessary to convince people of truth	HR	.10						−								
BPI	94	As a child belonged to a cohesive crowd or gang	ANX	.05		−						−						
BPI	401	Am entirely self-confident	SC	.10		−					−							
BPI	109	When with women, men usually think of sex	HR	.01														
BPI	203	Mother had more brains and ability than father	HR, SC	.01, .10														
BPI	134	More sensitive to odors than most people	SC	.10														
BPI	12	Not person who tries to "sell" himself	ANS	.05														
					Free or easy impulse expression	Social ease	Extroverted	Rebellious, independent, dominant	Secure	Active	Untroubled	Involved with friends	Trusting	Contented or socialized	Motoric	Cynical	Curious, creative, intellectual	Father centered
BPI	10	When bored like to stir up excitement	ANX	.10	+	+	+	+	+	+				−	+			
BPI	375	Often say humorous things	HR, ANX	.01, .10	+	+	+	+	+	+							+	
BPI	135	Person needs to show off	HR	.05	+	+	+	+	+	+								
BPI	418	Like to flirt	ANX	.05	+	+	+	+	+	+					+			
SEI	30	Enjoy wild parties	ANX	.10	+	+	+	+		+								
BPI	484	Like to be center of attention	HR, SC, ANX	.01, .05, .01	+	+	+	+		+								
MMPI	449	Enjoy social gatherings to be with people	HR	.05	+	+	+	+		+								
BPI	474	More enthusiastic than average	HR	.05	+	+	+			+		+	+			−	+	
BPI	374	Take part in entertainment at parties	HR, ANX	.05, .05	+	+	+		+	+		+						
BPI	408	Describe self as strong personality	ANX	.10	+	+		+	+	+								

Table 2—Continued

Source					Consensus rating of personality dimensions[a]													
Test	Item no.	Condensed text of item	Criterion[b]	Significance level	Free or easy impulse expression	Social ease	Extroverted	Rebellious, independent, dominant	Secure	Active	Untroubled	Involved with friends	Trusting	Contented or socialized	Motoric	Cynical	Curious, creative, intellectual	Father centered
MMPI	447	Go out of way to win a point	HR, ANX	.05, .05	+	+		+		+	–			–				
IES		Would like to drive racing car	SC, ANX	.05, .05	+		+	+		+					+			
BPI	394	Often act on spur of moment	ANX	.05	+		+	+		+					+			
BPI	62	Have good sense of rhythm	SC, ANX	.10, .05	+		+			+					+			
BPI	381	In school was sent to principal	HR	.10	+			+		+					+			
BPI	426	Do not blame opportunists	ANX	.10	+			+	–	+		–	–	–		+		
BPI	159	Life is to be senuously enjoyed	ANX	.01	+			+		+							+	
BPI	380	Like large, noisy parties	ANX	.10	+	–	+			+								
BPI	458	Like to let people know where I stand	HR, ANX	.05, .05	+			+		+								
IES		People would be happier if premarital sex were accepted custom	ANX	.05	+			+										
IES		Like to listen to primitive music	SC, ANX	.10, .10	+												+	
BPI	83	On trip, prefer talking to people to enjoying scenery	HR	.10		+	+					+	+	+			+	
BPI	375	Try to remember good stories to pass on	HR	.01		+	+					+		+				
MMPI	107	Am happy most of the time	HR	.05		+			+		+		+	+				
BPI	263	Have talent for influencing people	ANX	.10		+		+	+	+								
BPI	333	Am pretty poised person	ANX	.01		+		+	+									
BPI	236	Enjoy giving advice and help	ANX	.05		+				+		+						–
BPI	147	Go out of way to see what crowd is gathered for	HR	.05		+				+								
BPI	166	Know how others expect me to behave	ANX	.05		+												
MMPI	146	Not happy unless traveling	ANX	.10				+		+		–		–	+		+	
MMPI	505	Had long periods of pep when sleep seemed unnecessary	HR	.01				+		+					+			
MIS		More like father	HR	.10				+		+								+
BPI	448	More curious than most	HR	.05				+									+	
IES		Enjoy discarding old and accepting new	ANX	.05				+									+	
BPI	177	At times checked on excuses or stories	ANX	.10				+									+	
IES		Problem attracts me until I can find answer	HR	.10				+							–		+	
IES		Attention should be paid to new ideas, even if radical	ANX	.10				+							–		+	
BPI	207	Do best work at last minute	ANX	.10				+										
BPI	391	Do best work late at night	ANX	.10				+										
MMPI	137	Home life as pleasant as most people's	ANX	.05					+		+			+				
BPI	251	Members of family very close	ANX	.05					+			+		+				
BPI	65	Categorize self as optimist	HR	.05					+				+	+				
BPI	77	Not sick or frightened by blood	ANX	.05					+		+							
MMPI	155	Neither gaining nor losing weight	SC	.01							+							
MMPI	103	No trouble with muscles twitching	ANX	.05							+							
IES		Eyesight as good as has been	ANX	.05							+							
BPI	358	Speech is same as always	ANX	.10							+							
BPI	366	No more nervous than most	ANX	.10							+							

Table 2—Continued

Source		Condensed text of item	Criterion[b]	Significance level	Consensus rating of personality dimensions[a]													
Test	Item no.				Free or easy impulse expression	Social ease	Extroverted	Rebellious, independent, dominant	Secure	Active	Untroubled	Involved with friends	Trusting	Contented or socialized	Motoric	Cynical	Curious, creative, intellectual	Father centered
BPI	198	Prefer old friends to potential new ones	SC	.05		−		−	−			+		+				
MMPI	317	More sensitive than most	SC	.10								+						
BPI	152	Get upset about friend's misfortune	HR	.05							−	+						
MMPI	91	Do not mind being made fun of	ANX	.10				−		−			+					
BPI	353	Have good deal of self-control	ANX	.10	−									+				
BPI	23	Aware of feelings, changes, activities in body	ANX	.10											+			
BPI	63	Wise person sees life as game, watches and wonders	ANX	.10										−		+	+	
BPI	432	Average person not able to appreciate art and music	HR, ANX	.05, .10										−		+		
MMPI	124	Most people will use somewhat unfair means	ANX	.05									−			+		
BPI	79	Person with sense of responsibility is taken advantage of	ANX	.05					−				−			+		
BPI	96	Do not respect most people I know	ANX	.05								−	−	−		+		
BPI	42	People are reluctant to exert themselves for others	ANX	.05		−			−		−	−	−	−		+		
BPI	111	Read 10 books a year	ANX	.05													+	
BPI	100	Capable of deed benefiting the world	ANX	.10													+	
BPI	321	Have fun making limericks, word plays, or rhymes	HR	.05											−		+	
BPI	161	Like to work out puzzles	HR	.10			−										+	
BPI	255	Would like to be like father	HR	.05														+
BPI	15	Prefer to work with women	SC, ANX	.10, .10														−
BPI	331	Consider all sides before making decision	SC	.10											−			
BPI	140	A discreet woman can do as she pleases	ANX	.05										−				
BPI	270	Many people might answer questionnaire untruthfully	ANX	.10									−					
BPI	142	Would reject friend of repulsive appearance even if brilliant	SC	.10								−					−	
BPI	202	Reluctant to lend books, clothes, etc.	ANX	.10					−		−		−					
BPI	346	Rember playing sick to get out of something	HR	.05				−		−								
MMPI	431	Worry over possible misfortunes	SC	.10				−	−		−		−					
MMPI	278	Strangers have looked at me critically	ANX	.10		−		−	−		−		−					
BPI	267	Parents always strict and stern	ANX	.05	−				−		−							
BPI	212	Often avoid making scene by giving in	ANX	.10	−			−	−	−							−	
BPI	310	Can imagine what father like at 20	HR	.05														
BPI	26	Enjoy excitement of a crowd	HR	.10														
APQ		When anxious get sinking feeling in stomach	HR	.10														
BPI	370	Good to know people who can fix tickets, etc., for you	SC, ANX	.05, .05														
MMPI	6	Like to read articles on crime	SC	.10														
MMPI	178	Memory seems all right	HR	.05					+		+							

DISCUSSION

Earlier, theoretical arguments were presented to explain why ipsative experimental designs should be considerably more sensitive than normative designs in detecting personality correlates of styles of stress response. Although any behavior or response may be related to personality, experimenters' hypotheses about personality correlates usually predict not differential behavior in all situations, but different behavior in different situations. Hence, comparison of responses of the same individuals over different situations is usually the experimental design of choice. Our empirical findings are certainly consistent with this theoretical analysis. When the ipsative approach was used, many more relationships appeared between the personality-questionnaire items and stress-response measures. Moreover, the relationships made sense in that the persons who responded more to one stress condition than to the other described themselves in an internally coherent and plausible manner. Considerable consistency in these relationships was observed over three quite separate stress-response modalities, heart rate, skin conductance, and anxiety.

It would be premature to interpret the particular personality relationships found in this study. Without replication it is not possible to know which of the discriminating questionnaire items would stand up under cross-validation. Nor it it known which of the several dimensions of difference between the two stress conditions was crucial. Further experiments are necessary to replicate the substantive findings and to evaluate the process by which the personality variables are linked to the different stress conditions. By separating out the psychological components in each of the complex stress conditions, we can hope to evaluate their contribution to the obtained relationships. We can compare, for example, a direct threat of shock with a vicarious threat of shock; likewise, we can compare a vicarious social threat with a direct social threat, a direct shock with a direct social threat. Such an experiment is currently in progress, and, while the results may alter present speculations about the attributes of personality that mediate the selective reactions observed, they cannot weaken the logical arguments presented for an ipsative, differential approach to the relationship between personality and stress response.

REFERENCES

Aas, A. The measurement of subjective experience presumably related to hypnotic susceptibility. *Scandinavian Journal of Psychology,* 1962, 3, 47–64.

Alfert, E. Comparison of responses to a vicarious and a direct threat. *Journal of Experimental Research in Personality,* 1966, **1**, 179–186.

Benjamin, L. S. Statistical treatment of the law of initial values (LIV) in autonomic research: A review and recommendation. *Psychosomatic Medicine,* 1963, **25**, 556–566.

Block, J. A study of affective responsiveness in a lie-detection situation. *Journal of Abnormal and Social Psychology,* 1957, **55**, 11–15.

Block, J. On the number of significant findings to be expected by chance. *Psychometrika,* 1960, **25**, 369–380.

Craig, D. R. An investigation of basic skin resistance levels during sleep under differing conditions. Unpublished master's thesis, Indiana University, 1936.

Darling, R. P. Autonomic action in relation to personality traits of children. *Journal of Abnormal and Social Psychology,* 1940, **35**, 246–260.

Darrow, C. W. Sensory, secretory and electrical changes in the skin following bodily excitation. *Journal of Experimental Psychology,* 1927, **10**, 197–226.

Engel, B. T. Stimulus-response and individual response specificity. *Archives of General Psychiatry,* 1960, **2**, 305–313.

Eysenck, S. B. G. An experimental study of psychogalvanic reflex responses of normal, neurotic, and psychotic subjects. *Journal of Psychosomatic Research,* 1956, **1**, 258–272.

Grossman, H. J., & Greenberg, H. H. Psychosomatic differentiation in infancy: I. Autonomic activity in the newborn. *Psychosomatic Medicine,* 1957, **19**, 293–306.

Jones, H. E. The study of patterns of emotional expression. In M. L. Reymert (Ed.), *Feelings and emotions: The Mooseheart Symposium.* New York: McGraw-Hill, 1950. Pp. 161–168.

Jost, H., & Sontag, L. W. The genetic factor in autonomic nervous-system function. *Psychosomatic Medicine,* 1944, **6**, 308–310.

Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. *Annals of the New York Academy of Sciences,* 1956, **67**, 123–164.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy.* Washington, D. C.: American Psychological Association, 1959. Pp. 160–208.

Lacey, J. I., Bateman, D. E., & Van Lehn, R. Autonomic response specificity and Rorschach color responses. *Psychosomatic Medicine*, 1952, **14**, 256–260.

LACEY, J. I., KAGAN, J., LACEY, B., & MOSS, H. A. The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), *Expression of the emotions in man*. New York: International Universities Press, 1963. Pp. 161–196.

LEHTOVAARA, A., SARRINEN, P., & JÄRVINEN, J. Psychological studies on twins: I. GSR reactions: Environmental versus hereditary determination. In, *Reports from the Psychological Institute*. Helsinki: University of Helsinki, 1965.

LYNN, R., & EYSENCK, H. J. Some effects of Carisprodol on pain reactivity. In H. J. Eysenck (Ed.), *Experiments with drugs*. New York: Macmillan, 1963. Pp. 324–328.

MANDLER, G., MANDLER, J. M., & UVILLER, E. T. Autonomic feedback: The perception of autonomic activity. *Journal of Abnormal and Social Psychology*, 1958, **56**, 367–373.

NOWLIS, V., & NOWLIS, H. The description and analysis of mood. *Annals of the New York Academy of Sciences*, 1956, **65**, 345–355.

OKEN, D., GRINKER, R. R., HEATH, H., HERZ, A., KORCHIN, S. J., SABSHIN, M., & SCHWARTZ, N. The relation of physiological response to affect expression. *Archives of General Psychiatry*, 1962, **6**, 336–351.

PAVLOV, I. P. *Conditioned reflexes*. Oxford: Oxford University Press, 1927.

RICHMOND, J. B., & LUSTMAN, S. L. Autonomic function in the neonate: I. Implications for psychosomatic theory. *Psychosomatic Medicine*, 1955, **17**, 269–275.

ROESSLER, R., ALEXANDER, A. A., & GREENFIELD, N. S. Ego strength and physiological responsivity. *Archives of General Psychiatry*, 1963, **8**, 142–154.

SANFORD, N., WEBSTER, H., & FREEDMAN, M. Impulse expression as a variable of personality. *Psychological Monographs*, 1957, **71**(11, Whole No. 440).

SEYMOUR, R. B. Personality correlates of electrodermal resistance and response. Unpublished doctoral dissertation, University of California, Berkeley, 1950.

STERNBACH, R. A., & TURSKY, B. Ethnic differences among housewives in psychophysical and skin potential responses to electric shock. *Psychophysiology*, 1965, **1**, 241–246.

TITELBAUM, S. The electrical skin resistance during sleep. Unpublished doctoral dissertation, University of Chicago, 1941.

VOGEL, W., RAYMOND, S., & LAZARUS, R. S. Intrinsic motivation and psychological stress. *Journal of Abnormal and Social Psychology*, 1959, **58**, 225–233.

ZBOROWSKI, M. Cultural components in response to pain. *Journal of Social Issues*, 1952, **8**, 16–30.

(Received July 25, 1966)

Manuscripts Accepted for Publication in the *Journal of Personality and Social Psychology*

Effectance Arousal and Attraction: Donn Byrne* and Gerald L. Clore, Jr.: Department of Psychology, the University of Texas, Austin, Texas 78712.

Trait Inferences: Evaluative and Descriptive Aspects: Dean Peabody*: Department of Psychology and Education, Swarthmore College, Swarthmore, Pennsylvania 19081.

Use of Role Reversal in Intergroup Competition: David W. Johnson*: Department of Educational Psychology, University of Minnesota, Minneapolis, Minnesota 55455.

Social Anxiety, Opinion Structure, and Opinion Change: David O. Sears*: Department of Psychology, University of California, Los Angeles, California 90024.

Modeling Effects on Achievement Imagery: Eric Klinger*: Division of the Social Sciences, University of Minnesota, Morris, Minnesota 56267.

Effects of Parental Status Incongruence and Discrepancy on Physical and Mental Health of Adult Offspring: Stanislav V. Kasl*: Institute for Social Research, The University of Michigan, Ann Arbor, Michigan 48104.

Ethnocentric Dogmatism and Binocular Fusion of Sexually and Racially Discrepant Stimuli: Marvin A. Iverson* and Helen G. Schwab: 154 Charles Street, East Williston, New York 11596.

Personality Correlates and Situational Determinants of Rate of Change in Attractiveness of Success and Repulsiveness of Failure in Relation to Task Difficulty: N. T. Feather*: Department of Psychology, The University of Michigan, Ann Arbor, Michigan 48104.

Involvement as a Determinant of Response to Favorable and Unfavorable Information: Alice H. Eagly*: Department of Psychology, Michigan State University, East Lansing, Michigan.

Cognitive Factors in the Recognition of Ambiguous Auditory and Visual Stimuli: John R. Frederiksen*: Center for Cognitive Studies, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

Reactions of Negroes in a Biracial Bargaining Situation: John M. Hatton*: Counseling and Testing Center, Stanford University, Stanford, California.

Choosing to Suffer as a Consequence of Expecting to Suffer: A Replication: Frank Rosekrans*: 494 Ford Hall, Laboratory for Research in Social Relations, University of Minnesota, Minneapolis, Minnesota 55455.

Imatative Influences of Consistent and Inconsistent Response to Consequences to a Model on Aggresive Behavior in Children: Mary A. Rosekrans* and Willard W. Hartup: Institute of Child Development, University of Minnesota, Minneapolis, Minnesota.

---

* Asterisk indicates author for whom address is supplied.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 304-312

BILINGUAL SEMANTIC MERGING AND AN ASPECT OF ACCULTURATION

MARGARET J. EARLE[1]

University of Hong Kong

Chinese-English bilingual students were classified according to the degree of assumed similarity between a sample of their own beliefs and those of representatives of the native-language community, and the 2nd-language community. Ss more similar to the 2nd-language community obtained significantly lower semantic differential D scores between Chinese and English translation-equivalent concepts than did Ss more similar to the native-language community. Students attending an English-language university obtained significantly lower D scores than students attending an English-language secondary school. Results are related to the theory of compound and coordinate bilingualism, and it is suggested that acceptance of beliefs of the 2nd-language community may result in a transformation from a coordinate to a compound bilingual system.

The nature of the semantic relationship between the two languages of the bilingual person raises questions which have received attention from both linguists and psychologists. On the basis of linguistic sign theory Weinreich (1953) has suggested that there are three possible kinds of relationship between the two languages of the bilingual speaker. If he interprets translation-equivalent words as having different meanings, they are considered to belong to coexistent systems and are two distinct signs. If the two words have the same meaning for the individual, they form a compound sign and belong to partially merged systems. If one word refers only to a word in the other language and not to a nonlanguage event, they belong to completely merged systems. Ervin and Osgood (1965) have proposed a distinction between two types of meaning relationship on the basis of mediation theory and the process of learning the second language. The *compound* bilingual is said to have learned the second language by associating each new word with a word in the native language. He has a common set of representational mediation processes (meanings) for translation-equivalent words in the two languages. The *coordinate* bilingual is described as having learned to speak and use the two languages in different external and emotional situations, as, for example, one language at home and the other at school and work. The representational mediation processes with which words in one language are associated are somewhat different from those with which words of the second language are associated.

Several studies of French-English bilinguals in Canada have been designed on the basis of the compound and coordinate theory; three behavioral indexes have been reported to differentiate between the two types of bilinguals. Jacobovits and Lambert (1961) found a semantic satiation effect across the two languages of compound bilinguals, but no such effect with coordinate bilinguals. Lambert, Havelka, and Crosby (1958) reported that experience in separated language-acquisition contexts (coordinate learning) resulted in greater associative independence of translation equivalents in the two languages. Separated subjects were unaffected by the learning of a list of French words interpolated between the learning and relearning of the translation-equivalent English words. On the contrary, subjects who had acquired their two languages in fused contexts (compound learning) benefited from the interpolated French list.

Results obtained by Lambert et al., using the semantic differential *D*-score measure (Osgood, Suci, & Tannenbaum, 1957) showed also that the difference in connotative meaning between translation-equivalent words in French and English is greater for those who learned the two languages in separated contexts than for those who learned in fused contexts. However, Lambert et al. distinguished between two groups of separated subjects: those who learned both languages in the same cultural or national setting (unicultural), and those who learned the two languages in different cultural settings (bicultural). The authors found that the mean *D* score of separated unicultural subjects did not differ significantly from the mean *D* score of fused subjects, although scores of the latter group were lower. The mean *D* score of the separated bicultural subjects was significantly higher than that of both the other two groups. They offered the explanation that in the bicultural learning situation there is less likelihood of the actual referents of words being the same. This suggests that, although the unicultural learning situation may produce separated (or coordinate) bilinguals, it also provides learning experience more similar to that of the compound bilingual as described by Ervin and Osgood.

---

[1] Now at the University of Aberdeen.

The compound and coordinate classification is based upon a theoretical discussion of what happens when the second language is being learned. The studies described above have categorized subjects in terms of this learning process. It seems unlikely that the semantic relationship between the two language systems would always remain static, as determined by the initial learning experience. Haugen (1956) referred to the second category in Weinreich's tripartite classification as appearing to be somewhere between the pure compound system and the pure coordinate system as defined by Ervin and Osgood. Haugen continued that this would be "typical of the immigrant who has been socially assimilated but who continues to speak his native language: most of the meanings he now associates with its morphemes are those of the dominant culture of his new homeland [p. 79]." This statement implied that the meaning realtionship between the bilingual's two languages will change with social assimilation. Although Haugen did not phrase this in psychological terms, a question is raised for the psychologist as to the possible effects of assimilation on the meaning systems of the bilingual.

Assimilation has been described as a phase of acculturation. Beals (1953) interpreted this to mean "that assimilation is that form of acculturation which results in groups of individuals wholly replacing their original culture by another [p. 628]." Beals has discussed in detail the ways in which the concept of acculturation is used by anthropologists. From the psychologist's point of view the process and results of acculturation may be seen in terms of the effects on the individual. One aspect of the results of acculturation can be formulated, following Beals, as acceptance of behavior patterns and values of the new culture. It would seem reasonable, as a preliminary definition of acculturation of the individual, to suggest that he accepts the beliefs, values, and customs of a new culture. One can then argue that his acceptance would have an effect on the connotative meanings of an acculturated bilingual, or a bilingual who is undergoing a process of acculturation.

Meaning, as measured by the semantic differential, would seem to be rather closely related to beliefs and values associated with particular concepts. For example, if the concept TREE is nice, strong, and clean in French, one would expect certain associated beliefs about the aesthetic and functional value of trees. If, in English, TREE is nasty, weak, and dirty, quite different beliefs would be expected. If the French speaker comes to understand, accept, and adopt an English speaker's view of trees, it would be expected that his ratings on the semantic differential would change somewhat towards the ratings given by the English speaker. To the extent that the French speaker makes the English beliefs his own, this change should affect his ratings in French as well as in English. In Haugen's terms, the meanings associated with the morphemes of the native language would become those of the new culture, or in this case, an influential representative of that culture. If we assume that aculturation involves acceptance of new beliefs and modifications of old beliefs, then it might also involve some modification of old connotative meanings. These considerations lead to the hypothesis that increased contact with and acceptance of the second-language culture would result in a closer relationship between connotative meanings in the two languages of the bilingual, even though the second language was acquired in a separated context.

Hong Kong provides a situation where native Chinese speakers learn and use English in a separated context, and many Hong Kong bilinguals have learned both their Chinese and their English in Hong Kong. Furthermore, some of these bilinguals receive a secondary-school and university education in their second language and are educated in the ideas and technology of the second-language community. They are, through their education, as well as through day-to-day living in a British-administered and Western-influenced society, exposed to many contacts with the second-language culture. There is considerable evidence of a general impressionistic nature that this exposure leads some of the students to accept certain aspects of the second-language culture. However, systematic studies of acculturation have not been done in Hong Kong, and there is very little information available as to the nature, direction, and degree of belief change which might take place in this two-culture environment. Development of a direct measure of acceptance of the second-language culture would have to await extensive background research.

In designing the present exploratory experiment an indirect measure was adopted, based on certain assumptions about belief acceptance in this particular social context. It would be convenient to ask the bilingual himself to indicate similarities and differences between certain of his own beliefs and those which he attributes to members of each of the two language communities. This would provide a measure of assumed similarity between himself and members of his two language communities. Assumed similarity may mean either actual similarity or a desire to be similar, and the latter does not necessarily imply acceptance. However, it is assumed that a desire to be similar would be a motive for acceptance. When the individual has spent his life in a society greatly influenced by the second-language community, and has had about 10 years of formal education from the second-language community, it seems reasonable to suppose that a desire to be similar would have led him to some acceptance. It was therefore presumed that assumed similarity in this context does imply acceptance, and that the individual can approximately define his acceptance by his assumed-similarity responses. On the basis of these assumptions the hypothesis was reformulated as follows:

If Chinese-English bilingual subjects see their own beliefs as being more similar to the beliefs of the second-language community than to those of the native-language community, their ratings on semantic differentials should be closer in English and Chinese (lower D) than would be the ratings of subjects who see their own beliefs as more similar to the native-language community.

EXPERIMENT I

Method

Subjects. Thirty-six first-year students at the University of Hong Kong served as subjects in the first section of the study.[2] There were 10 males and 26 females whose average age was 19 years. All were native Cantonese speakers. All the subjects had first learned English in school as a second language beginning at the age of 6 or 7; all had received their secondary (high school) education at Anglo-Chinese schools in Hong Kong, where the medium of instructions is English. From the information available, all the subjects appear to be in the coordinate bilingual category. Cantonese is used at home almost exclusively and is spoken in the University by the students among themselves. English is the language of formal education and has been for all subjects for approximately 8-9 years. It is not possible to say with certainty that these students are coordinate bilinguals, but it seems clear that the native and second languages have been acquired and used in almost entirely separated contexts. Furthermore, in a repetition of the Jacobovits and Lambert (1961) experiment on semantic satiation, it was found that these and other students at the University of Hong Kong do not show any semantic satiation effect across their two languages, a result which Jacobovits and Lambert report for their coordinate subjects.

Procedure

Measurement of connotative meaning. Semantic differential scales in English and Chinese were used to measure the connotative meaning of 12 concepts in English and Chinese translation equivalents. The procedure followed was similar to that of Lambert et al. but with one important difference. These authors used translations of the English semantic differential for their French version of the scales.

---

[2] Initially there were 43 students in the group, but 7 failed to complete the semantic differentials.

In the present study the English semantic differential scales used were those developed in America, and the Chinese scales were those developed in Hong Kong in Cantonese.[3] Scales selected were from those having the highest loadings on the three principal factors of the concept on scale factorizations (Osgood, 1964). Four scales were used for each factor, giving a total of 12 scales in English and 12 in Chinese. Thus the Chinese semantic differential was not a translation of the English.

The concepts used were selected from the list of 100 stimulus words used in the cross-cultural studies of semantic differentials (Osgood, 1964). Concepts were selected from this list because considerable care had been given to translation by a panel of Hong Kong bilinguals, and hence translation problems had been largely solved. Concepts used were FOOD, BOOK, FATHER, MOTHER, WORK, RESPECT, FREEDOM, KNOWLEDGE, FRIEND, FUTURE, SUCCESS, PLEASURE, and Chinese translations. The concepts were presented in a fixed order so that each one was as far away as possible from its translation. *D* scores were calculated for each subject, using the formula given by Osgood et al. (1957).

Similarity of beliefs with the second-language community. The method devised to assess the extent of belief similarity was an adaptation of the response-prediction technique, as used in studies of certain aspects of identification; a measure of assumed similarity (Bronfenbrenner, 1958; Marwell, 1964). Twenty statements of beliefs were prepared in English.[4] These were intended to cover some of the main areas in which Chinese students might consider that their beliefs differ from those of English-speaking Western students. The statements were based upon material derived from discussions with University of Hong Kong students and graduates and were concerned with beliefs about authority, independence, traditions, and the use of leisure.

Subjects were asked to indicate the degree of agreement or disagreement with each statement on a 6-point scale, from "strongly agree" to "strongly disagree." Four forms of the statements were given to each subject, the first form with instructions to give their own personal opinions. The other three forms were distributed in systematically varied order. Instructions were to (*a*) answer as the typical British student would, (*b*) answer as most other Chinese students would, (*c*) answer as the typical American student would. The instructions made it clear that subjects were not expected to know how anybody else would answer the questions, but they were asked to guess on the basis of their general knowledge of the other groups.

Results were calculated by assigning a weighting to each point on the scale from most strongly agreeing (7) to most strongly disagreeing (1). The subjects' personal scores were taken as the standard. Discrepancy scores were calculated by subtracting scores for each statement, that is, between self and Chinese, self and English, and self and American. The discrepancy scores for each subject were then added, given a total discrepancy score over 20 statements between self and each of the three language-culture groups, with a possible range of 0–120.

Subjects were given the semantic differentials first, followed by the belief questionnaire.

Results

Analysis of belief discrepancies.[5] The analysis of belief ratings was designed to identify those subjects who regard themselves as more similar to Chinese and those who regard themselves as more similar to British and American. In all cases the self-British and self-American discrepancies were similar, making it possible to combine the two into a general English-speaking category, taking the average of the two discrepancy scores for each subject. Classification of subjects as closer to Chinese or English speaking was necessarily determined by an arbitary standard. This was based upon the mean difference between self–Chinese (SC) and self–English-speaking (SE) discrepancies. Subjects were assigned to three categories, called, for convenience, Chinese affiliate, English affiliate, and no differences. The classification was made in the following way:

1. The mean difference between SC and SE was calculated for the group ($M = 21.05$).
2. The Chinese affiliate group includes all subjects whose SC discrepancy was at least 21 points lower than the SE discrepancy.
3. The English affiliate group includes all subjects whose SE discrepancy was at least 21 points lower than the SC discrepancy.
4. All other subjects were assigned to the no difference group.

Mean discrepancy scores for each of the three groups are shown in Table 1.

Relationship between belief affiliation and semantic differential D *scores*. Mean semantic differential *D* scores were calculated for each of the three affiliation groups, giving the dif-

---

[3] The author is grateful to Anita K. Li, Department of Education, University of Hong Kong, for providing the Chinese semantic differential material.

[4] Although Cantonese is the first language of these subjects, most of them have had more experience in reading English than reading Chinese.

[5] It should be noted that, although this study uses only rating techniques, the method used to classify acceptance of beliefs minimizes the likelihood of results being artifacts of response set.

TABLE 1

BELIEF AFFILIATION SCORES: MEAN DISCREPANCY SCORES FOR THREE AFFILIATION GROUPS

	Mean SC discrepancy	Mean SE discrepancy
Chinese affiliate ($N = 12$)	9.00	32.00
English affiliate ($N = 15$)	33.87	9.07
No difference ($N = 9$)	18.00	19.60

ference in connotative meaning between the 12 concepts in English and Chinese. The hypothesis predicts that the English affiliate group will have significantly lower D scores than the Chinese affiliate subjects. This is supported by the results shown in Table 2.

These results do not give any indication as to the cause-effect relationship between belief affiliation and concept meanings. The introductory discussion suggests that a move towards acceptance of the beliefs of the second-language community results in a closer semantic relationship between the two languages. The causal implications of this could, ideally, only be tested by a longitudinal study of the same bilinguals over a number of years. The present investigation was designed as no more than a preliminary exploration of the problem, and no such longitudinal study has yet been undertaken. It is, however, possible to obtain some indirect evidence as to the cause-effect relationship by taking another cross-section of Hong Kong bilinguals, who are at a somewhat earlier stage of their English education, and who might be presumed to be both less familiar with, and less committed to, the beliefs of the second-language community. Such a group of subjects was found in an Anglo-Chinese school which, as far as possible, is representative of the schools from which the University students have come. A second hypothesis was therefore tested, based upon the assumption that students in secondary school are likely to be less familiar with the beliefs of the second-language community, having had a shorter period of British education at a lower level; and are less likely to have accepted these beliefs, having not yet committed themselves to at least 3 years of higher education at a British university, and perhaps not intending to do so.

TABLE 2

MEAN SEMANTIC DIFFERENTIAL D SCORES FOR THE THREE AFFILIATION GROUPS: UNIVERSITY STUDENTS

	Mean D score	SD
Chinese affiliate ($N = 12$)	6.45	1.14**
English affiliate ($N = 15$)	5.19	.88
No difference ($N = 9$)	5.67	1.13

** $p < .01$, $t = 3.3$ (1-tailed test).

The hypothesis tested in the second part of the study was formulated as follows:

Students in an Anglo-Chinese secondary school will differ from students in the University of Hong Kong in that fewer of them will fall into the English affiliate group, and their D scores on the semantic differential will be significantly higher.

EXPERIMENT II

Method

Subjects. It was decided, after discussion with school teachers, that the methods used would be rather too complicated for students below the sixth form (last 2 years of high school) to understand. It was therefore necessary to take a group of subjects who were only from 1 to 3 years behind the University students in the educational system. The subjects were 36 students, 18 males and 18 females, at an Anglo-Chinese school which prepares students for admission to the University of Hong Kong.[6] Twenty-five of the subjects were in the lower sixth form and 11 in the upper sixth form. The average age was 17½ years. All were native Cantonese speaking, and, although not as much information was available as for the University subjects, it would seem that the language experience of the two groups had been very similar, except for the additional few years of English which the University students have had.

---

[6] Of the initial 58 students in the group, 22 failed to complete either the semantic differentials or the belief ratings and had to be eliminated. It is coincidental that the number of subjects remaining was the same as the number in the University group. The high attrition rate appears to be due in part to the failure of some students to understand the task, but largely to shortage of time. The experiment had to be fitted into the school timetable.

TABLE 3

BELIEF AFFILIATION SCORES: MEAN DISCREPANCY SCORES FOR THREE AFFILIATION GROUPS—SECONDARY-SCHOOL SUBJECTS

	Mean SC discrepancy	Mean SE discrepancy
Chinese affiliate ($N = 11$)	8.00	39.50
English affiliate ($N = 9$)	42.16	12.50
No difference ($N = 16$)	22.80	21.00

Procedure

The methods used were the same as those in Experiment I, but because of time limitation only six concepts were used for semantic differentials: FOOD, BOOK, FATHER, MOTHER, WORK, RESPECT, and Chinese translations. Analysis of results was the same as for Experiment I.

Results

Subjects were classified into three belief affiliation groups as in Experiment I.[7] The mean discrepancy scores are shown in Table 3.

These results give some support to the first part of the hypothesis, that fewer of the secondary-school subjects would be classified as English affiliate, although with the small numbers involved it is difficult to draw any conclusion from this. It would be misleading to make statistical tests of these data, given the approximate nature of the measure of affiliation. However, it can be said that 41.6% of the University students were classified as English affiliate, as compared to 25% of the secondary-school students.

Relationship between belief affiliation and semantic differential D scores. Results from the secondary-school subjects support those of Part I of the study (see Table 4). *D* scores for the English affiliate group are again significantly lower than those for the Chinese affiliate group. The relationship between affiliation as defined here and *D* as measured by the semantic differential is thus found in two different samples. From the secondary-school subjects another result of some interest appears, namely, the position of the no-difference group. A relatively large number of subjects were classified into this group, and the mean *D* score was significantly lower than that of the Chinese affiliate group but the same as that of the English affiliate group. Not enough information is available to account for this result. One possible explanation which could be tested in the future is that some of these subjects are in a transitional period with regard to their allegiance to the two language cultures. It could be that, at this point in their educational experience, they are moving towards the second-language culture but have not yet fully recognized this.

TABLE 4

MEAN SEMANTIC DIFFERENTIAL *D* SCORES FOR THE THREE AFFILIATION GROUPS: SECONDARY-SCHOOL STUDENTS

	Mean *D* score	*SD*
Chinese affiliate ($N = 11$)	11.12	2.2[a]*
English affiliate ($N = 9$)	9.08	1.9[b]*
No difference ($N = 16$)	9.06	2.4

[a] $t = 2.12$.
[b] $t = 2.06$.
* $p < .05$ (1-tailed test).

Comparison of University and secondary-school subjects. As can be seen from a comparison of Tables 2 and 4, the *D* scores of the secondary-school students are higher than those of the University subjects. Scores for all subjects in each sample were compared, and the results are shown in Table 5. Mean *D* scores of the University subjects were calculated both for all 12 concepts and also

TABLE 5

COMPARISON OF *D* SCORES FOR UNIVERSITY AND SECONDARY-SCHOOL SUBJECTS

	Concepts	Mean *D* score	*SD*
University ($N = 36$)	12	5.76	1.25
University ($N = 36$)	6	5.44	1.5[a]*
School ($N = 36$)	6	9.58	2.39[b]*

[a] $t = 2.004$.
[b] $t = 2.07$.
* $p < .05$ (1-tailed test).

---

[7] The mean difference between SC and SE, used as the basis of classification, was 28.63 for the secondary-school subjects.

the 6 concepts used in the secondary-school study. In both cases the mean secondary-school score is significantly higher than the mean University score.

DISCUSSION AND CONCLUSIONS

Although the major results of this study support the hypothesis, this has been only a preliminary investigation of a rather broadly defined problem, and there are several points which require further clarification and, further, more precise investigation. Three interrelated questions arise, from which specific hypotheses for future research should be formulated.

1. The results of the two experiments show a relationship between assumed similarity of beliefs and closeness of connotative meanings in two languages. In designing the experiment, assumed similarity was taken to imply acceptance, in which case the results would show a relationship between acceptance of second-language community beliefs and closer connotative meanings. If, however, assumed similarity implies not acceptance but merely a desire to be similar or to appear to be similar, the results would have to be interpreted differently. The English affiliate subjects would be demonstrating only an unidentified motive to rate their own beliefs and those of the second-language community as similar. It would therefore be possible that the low *D* scores of these subjects represent not a measure of real similarity between meanings in the two languages, but merely the extent to which the subjects wanted the meanings to be similar. Thus the results of the experiment could be interpreted as due to a common factor, desire to be or appear to be similar, determining both relief ratings and semantic differential ratings.

With the information available it is not possible to discount this alternative interpretation. However, for the purpose of an exploratory investigation it is possible to consider which of the two interpretations is more likely to be correct. This can be done by examining the assumptions made above to link assumed similarity with acceptance, and also by considering the meaning of the *D* scores. It was argued above that a desire to be similar would result in acceptance if the opportunity is available to accept, that is, if the individual has sufficient information about members of the culture to whom he wants to be similar. The opportunity has certainly been available to the subjects of this experiment, both from general social contacts and particularly from formal education. If it can reasonably be argued that wanting to be similar is a motive for acceptance, then it would be difficult to account for continued failure to achieve a goal which is readily attainable. A desire to give the appearance of similarity poses a rather different problem which can only be considered here in very general terms by reference to some characteristics of Hong Kong society. Jarvie (1964) has discussed certain aspects of Westernization in Hong Kong and particularly a conflict between limited acceptance of the West and maintenance of the strong conviction that Chinese civilization is the greatest in the world:

> There are those few who accept the west as completely and as wholeheartedly as they can—mainly the educated teenagers, especially girls. The great majority of the society keeps the west at arm's length [p. 15].

From these observations it can be argued that subjects from this social and cultural background are unlikely to wish to demonstrate a false similarity with members of a culture regarded as inferior by the majority of their own people.

Nevertheless, the possibility remains that the alternative variable (desire for similarity) did influence results, and its effect on the semantic differentials must be considered. The low *D* scores of the English affiliate group may reflect false similarities between concept meanings. However, it is difficult in this case to account for subjects giving what can be called false similar responses. In answering the belief questionnaires the subjects were making comparisons in terms of specific beliefs between themselves and other groups. In responding to the semantic differentials they were rating a series of concepts (24 or 12) spaced so that the translation equivalents were maximally separated. They were given no indication that direct comparisons were to be made between the Chinese and

English scaling, and the semantic differentials were done before the belief ratings to eliminate any clues as to the nature of the task which might have been gained from belief comparisons. Subjects had had no previous experience with semantic differentials, and the questions asked after the experiment showed some bewilderment as to what had been going on. (This was taken to account in part for the high wastage rate of subjects particularly in the secondary school.) It therefore seems likely that subjects simply scaled the connotative meaning of each concept as it appeared. Since they were not presented with a Chinese-English comparison task, it is less likely that they attempted to create similarities which were false.

With doubt as to what was being measured by the belief ratings it is clearly impossible to draw any firm conclusions. However, it would seem that further research into the problem is justified by these results. More precise and direct methods are needed to discover what psychological changes might take place through close contact with the second-language community (as in higher education) which could affect the semantic systems of the bilingual.

Some additional information was available about the University students which suggests that belief acceptance as defined here may be related to Rokeach's (1960) concept of open-mindedness. Subjects in the English affiliate group obtained significantly lower scores on the Dogmatism scale (both in English and a Chinese translation) than subjects in the Chinese affiliate group. (I.e., the English affiliate subjects appear more open-minded.) Studies of the Dogmatism scale in Hong Kong have only recently been started, and the validity of the scale in Hong Kong has yet to be established. However, if the scale proves to be a valid measure of dogmatism in Hong Kong, the result mentioned here might open up a number of possibilities of defining more precisely cognitive aspects of belief acceptance.

2. The causal relationship between belief acceptance and semantic merging suggested here can only be adequately tested by a longitudinal study over a number of years. It could be argued that an initially closer relationship between meanings in the two languages enables the bilingual to accept the second-language community more readily. The results of the comparison between secondary-school and University students could be interpreted in this way. The question of cause and effect therefore remains open until a longitudinal study can be made.

3. There remains the question of how the results of this study might be interpreted in terms of the theoretical distinction between compound and coordinate bilingualism. The cause-effect assumption made here implies a shift from coordinate towards compound bilingualism under certain social and personal conditions. Ervin and Osgood's discussion of the process of two-way translation provides a clue as to how such a change might take place in coordinate bilinguals who are moving towards greater acceptance of the second-language culture. Translation for the coordinate bilingual is a relatively more complex process than for the compound bilingual. Ervin and Osgood (1965) predict that continued two-way translation would tend "to transform a coordinate system into a compound system [p. 145]."

A true coordinate bilingual maintains a complete separation between his two languages, "the total situations both external and emotional, and the total behaviors occurring when one language is being used will differ from those occurring with the other [p. 140]." In the case of the subjects in this study, complete separation will remain possible as long as the two languages and their cultures are kept in quite different contexts; English for formal British education and all contacts with English-speaking Hong Kong, and Cantonese for personal and social life which is Chinese. If, however, a student is being more profoundly influenced by his British education, to the extent of accepting and absorbing new beliefs, values, and customs, there will arise the problem of integrating his second-language ideas with those of the native-language culture. This problem is likely to be especially acute in a situation such as exists in Hong Kong, where the student is living in a largely Chinese community. He is continually engaged in communicating with members of both cultures, and, to the extent that he attempts to integrate the two, it would seem that he must become involved in two-way translation. The process may be generally described in the following way: A conversation in Cantonese with a monolingual enters an area where the bilingual has now adopted a new set of beliefs from his experience with English-speaking people. The topic is raised in Cantonese, the bilingual translates into English for his own benefit to consider his position, and then translates back into Cantonese for his reply. A similar process would take place in communication with a member of the English-language community if a topic of conversation involved interpreting ideas of the native-language culture. The most commonly quoted examples of this process in a Chinese context are conversations involving the subject matter in which a University graduate has specialized. A discussion of some aspect of human behavior in which special knowledge is relevant would result in the bilingual thinking in the language of psychology (English) while speaking in Cantonese. In the English-speaking context the most obvious example is a conversation which requires explaining a Chinese point of view which differs from that of the foreigner. Bilinguals report that in this case they would consider the Chinese point of view in Cantonese before explaining in English.

The bilingual who is not concerned with accepting and integrating the ideas of the second-language community would not be involved in this process to the same extent. First, special knowledge acquired in English remains confined to the content of that language and would not be relevant in a Cantonese discussion. Second, he would be less likely to engage in a conversation with an English-speaking person about the Chinese point of view and might consider that it would not be possible to explain anyway. In effect he maintains a complete separation between the two languages and all that is associated with them, and there is no need for translation. It is not suggested that such a person would never engage in two-way translation. The difference between the two types of bilingual outlined here is seen in terms of the frequency with which two-way translation is employed.

The translation process described provides a possible explanation of the present results which appears to be compatible with the compound and coordinate theory. This process, if it does occur, might also be relevant to the results reported by Lambert et al. The language experiences of the unicultural bilingual who has constant contacts with both language communities would merit further investigation.

REFERENCES

BEALS, R. Acculturation. In A. L. Kroeber (Ed.), *Anthropology today.* Chicago: University of Chicago Press, 1953. Pp. 621–641.

BRONFENBRENNER, U. The study of identification through interpersonal perception. In R. Tagiuri & C. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford: Stanford University Press, 1958. Pp. 110–130.

ERVIN, S. M., & OSGOOD, C. E. Second language learning and bilingualism. In C. E. Osgood & F. Sebeck (Eds.), *Psycholinguistics.* Bloomington: Indiana University Press, 1965. Pp. 139–146.

HAUGEN, B. E. *Bilingualism in the Americas.* Gainesville, Fla.: American Dialect Society, 1956.

JACOBOVITS, L., & LAMBERT, W. E. Semantic satiation among bilinguals. *Journal of Experimental Psychology*, 1961, **62**, 576–582.

JARVIE, I. C. Face and facade in Hong Kong. *New Society,* 1964, **3**, 13–15.

LAMBERT, W. E., HAVELKA, J., & CROSBY, C. The influence of language acquisition contexts on bilingualism. *Journal of Abnormal and Social Psychology,* 1958, **56**, 239–244.

MARWELL, G. Problems of operational definitions of "empathy," "identification" and related concepts. *Journal of Social Psychology,* 1964, **63**, 87–102.

OSGOOD, C. E. Semantic differential techniques in the comparative study of cultures. *American Anthropologist,* 1964, **66**, 171–200.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

WEINREICH, U. *Languages in contact.* New York: Linguistic Circle of New York, 1953.

(Received August 10, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 313-321

EFFECTS OF FEAR AND INSTRUCTIONS ON HOW TO COPE WITH DANGER [1]

HOWARD LEVENTHAL, JEAN C. WATTS, AND FRANCIA PAGANO

Yale University

Smokers were given fear-arousing communications on the dangers of smoking. 3 factors were manipulated: intensity of fear stimulus (moderate or high), receipt of instructions on how to stop smoking (given or not given), and smoking during the communication (Ss instructed to smoke or not to smoke while the dangers of smoking were being depicted). The high-fear communication strengthened desires to stop smoking, but had no effect on actual smoking behavior; the receipt of instructions on how to stop smoking had no effect on desires to stop smoking, but was highly effective in getting Ss to change their smoking behavior.

Four problems were investigated in the present study: (*a*) the relationship between fear arousal and acceptance of communications depicting a danger and recommending that subjects change their behavior so as to protect themselves from danger, (*b*) the relationship between verbal measures of acceptance taken immediately after the communications and behavior measures taken at later points in time, (*c*) the effects on acceptance of providing instructions on how to carry out the recommended change in behavior, and (*d*) the effects on acceptance of requiring subjects to perform the forbidden behavior while its dangerous consequences were being depicted in the communications.

Our approach to the first two problems was straightforward. Recommendations to stop smoking were presented along with fear-arousing communications on the danger of lung cancer from smoking, and subsequent changes in desire to stop smoking and in actual smoking behavior were examined. Previous studies have reported both positive and negative effects of highly fear-arousing antismoking communications on compliance with recommendations to stop smoking. Janis and Terwilliger (1962) found that a high-fear appeal produced slightly less attitude change against smoking than did a moderate appeal, and Leventhal and Niles (1964) reported that a mild control communication was more successful in strengthening intentions to stop smoking than either a moderate- or a high-fear appeal. However, a high-fear communication used by Niles (1964) was found to increase intentions to stop smoking among subjects who rated themselves as "highly vulnerable" to various health and safety hazards; among subjects rated as "low vulnerables," the high-fear appeal increased rather than decreased intentions to stop smoking. Complicating the picture still further is a study by Leventhal and Watts (1966) in which fear communications similar to those used by Leventhal and Niles (1964) failed to produce differences in intentions to decrease smoking, but had long-range effects on smoking behavior. In this study, subjects exposed to the high-fear communication were apparently more successful in decreasing smoking than were subjects exposed to the moderate- or low-fear appeal. It is possible that high-fear appeals create immediate resistance to compliance with protective recommendations but that in time this effect is reversed. But more data are needed to decide the point, since only in Leventhal and Watts (1966) were changes in smoking behavior measured, and in this study the measures were taken at only one point in time, 5 months after the communications were given.

The third and fourth problems require

---

[1] This investigation was presented by Francia Pagano in partial fulfillment of the Master of Public Health Degree at Yale University, 1965. The data were collected by the second and third authors during the latter's externship at the John Slade Ely Center for Health Education Research. The study was supported by the United States Public Health Service (CH 00077-04) and the United Health Foundation. The authors would like to thank Irvin Child and James Dabbs for critical comments respecting early drafts of the paper.

elaboration. It has been proposed (Leventhal, 1965) that the inconsistent effects of fear arousal on compliance with protective recommendations may be due to differences in the judged effectiveness of actions recommended to avert the alleged danger. An unintended effect of some high-fear communications may be to make subjects feel less able to act to avert the danger, and to arouse a greater need for reassurance that protective action is feasible and for instructions on how to carry out this action.

This possibility has been investigated in studies by Leventhal, Singer, and Jones (1965) and Leventhal, Jones, and Trembly (1966) in which fear-arousing communications on the danger of tetanus either included or withheld instructions on how to get a tetanus inoculation. All subjects (college students) were given the general recommendation to take a tetanus inoculation, but half of the subjects also received a map of the campus with the location of the student health service prominently marked and a brief review of typical daily activities which brought the student physically close to the health service. The instructions did not provide new information (students were already familiar with the location of the health service), but simply amounted to a plan for incorporating a visit to the health service into everyday activities much as one might plan to stop off at the bank or store on the way to work. In both studies it was found that a larger percentage of subjects who had received instructions took inoculations compared to those who did not receive instructions or received only the instructions without a preceding high- or low-fear appeal. However, the expectation that the instructions would be especially effective when subjects were exposed to a high-fear appeal was not confirmed; the percentage of high-fear subjects who took inoculations was equivalent to that of low-fear subjects, whether or not subjects received instructions.

For the present study a pamphlet giving specific instructions on how to stop smoking was prepared. It was expected that subjects who received these instructions would be more successful in decreasing smoking than subjects who did not receive these instructions, and that the instructions would be especially effective when subjects were exposed to a high-fear appeal. It was reasoned that if strong fear tends to make subjects feel helpless, while instructions help convince them that they can protect themselves, then the more difficult the protective action recommended, the more beneficial should be the effect of receiving instructions. Assuming that it requires more effort and persistence to stop smoking than to take a tetanus inoculation, then it is likely that instructions would be especially effective in persuading high-fear subjects to stop smoking, even though instructions failed to interact with fear appeal in persuading subjects to take tetanus inoculations (Leventhal et al., 1965, 1966).

The fourth problem concerned the effects of varying an association between the act of smoking and the dangerous consequences (poor health, surgery, risk of death from lung cancer) which allegedly result from smoking. It was thought that if some subjects were required to smoke during the danger communications and others instructed not to smoke, the association between response and danger would be stronger in the former condition and should lead to greater subsequent inhibition of smoking.

METHOD

Design

Three factors were manipulated: intensity of the fear stimulus (moderate or high), instructions on how to stop smoking (given or not given), and smoking during the communication (subjects instructed to smoke or not to smoke). These variables were combined with a subject factor, heavy smokers (subjects who smoked more than the sample median of 15 cigarettes per day) versus light smokers (subjects who smoked 15 or fewer cigarettes per day), to form a 2 × 2 × 2 × 2 factorial design. The analysis of smoking behavior included a fifth factor corresponding to the five time intervals for which reports of smoking behavior were obtained. A control group which received only the instructions on how to stop smoking was included to test the effect of receiving instructions without fear arousal.

Subjects

Subjects were 129 Yale students, two-thirds freshmen and one-third upperclassmen. Subjects were individually approached on campus by two male recruiters and asked whether they wished to participate for $1.50 in a program-evaluation study sponsored by the Public Health Service. Those wishing to participate (approximately 95% of those approached) then answered several questions on a card, including a question on present smoking habits. If the recruiter saw from the card that the volunteer was a cigarette smoker, he recruited him for the study. Subjects were then randomly assigned to the treatment conditions.

Subjects were given the experimental materials in a small room centrally located on campus. Eighteen groups (sessions) were run, two in each condition. Groups of from 6 to 8 subjects were originally scheduled for each session, but, since some subjects failed to keep their appointments and had to be rescheduled, the number of subjects at a session varied from 4 to 10 subjects. All subjects attending a given session received the same experimental materials. After being welcomed, subjects were reminded that the program was sponsored by the Public Health Service and that its purpose was to evaluate health-education materials. Subjects then completed two preexposure questionnaires, were given a communication,[2] and completed a postexposure questionnaire.

Smoking History and Attitudes

The first preexposure questionnaire included items dealing with the subjects' smoking history and current smoking habits. Also included were eight items dealing with (*a*) beliefs that smoking causes cancer, (*b*) judgments of personal vulnerability to lung cancer, (*c*) concern about the effects of smoking, (*d*) judgments of the effectiveness of decreasing smoking in preventing lung cancer, and (*e*) desire to quit and decrease smoking. All attitude questions in this and the following questionnaires were answered on 7-point scales with anchors at the midpoint and extremes (e.g., 1 = not at all, 4 = moderately, 7 = very much).

Mood Ratings

On a second preexposure questionnaire, subjects were asked to rate the degree to which each of 26 adjectives described their current moods or feelings. The items were related to the following dimensions: fear, depression, nausea, shame, vulnerability, aggression, helplessness. Many of these items were derived from the check list prepared by Nowlis (1965).

Smoking during the Film

When the first two questionnaires were completed, the experimenter told the subjects that they were to see a movie on smoking and said: "Contrary to what you might expect, we wish you to smoke during the film. We have provided cigarettes, matches, and ashtrays in case you have forgotten to bring your own." These items were then passed around. For those groups of subjects who were not to smoke, the following statement was substituted: "As you might expect, in view of the nature of the program, we should like to ask you not to smoke while you are here." Subjects in the smoking condition each smoked one cigarette. As this typically took between 7 and 8 minutes, subjects in the moderate-fear film smoked almost throughout their film, while subjects in the high-fear film usually had finished smoking before the surgery sequence was shown.

Fear Communications[3]

The moderate-fear appeal consisted of an 8-minute sound and color movie on the dangers of smoking. The movie demonstrated the way in which tars are drawn into the lung with cigarette smoke, paralyze the cilia of the bronchi, and damage the blood vessels and lungs. The relationship between smoking and lung cancer was elucidated by showing how tars are collected in the "throat" of a mechanical smoker, are trapped, chemically analyzed, and shown to contain known carcinogens. Scenes of invalids with respiratory troubles (a man out of breath after climbing three steps, an older man gasping for breath in a hospital bed) were included, and the film ended with the presentation of charts showing the relationship between cigarette sales and lung-cancer deaths, and between the probability of getting lung cancer and the number of cigarettes smoked. In brief, the film presented a relatively factual treatment of the dangers of smoking and stopped short of elaborating its most severe consequences.

The high-fear film consisted of the film described above followed by a 6-minute film showing a lung-cancer operation. The operation sequence showed the initial incision and forcing apart of the ribs, the removal of the black and diseased lung, the open cavity and beating heart, and finally the suturing of the wound. The film was in color and thoroughly authentic. In the more "gory" sequences many subjects were observed to squirm and lower their heads, and three subjects were sufficiently shaken to have to leave the room temporarily.

Instructions on How to Stop Smoking

Immediately after seeing the movie, subjects in the specific instructions condition read a pamphlet[4] giving detailed information on how to break the smoking habit. The reader was advised to make a list of reasons why he smoked and compare it to a second list of reasons why he might want to stop smoking. He was then asked to choose a "stop smoking day" and was given considerable advice on how to avoid conditions which encourage smoking (e.g., do not eat with friends who are smoking, prepare excuses beforehand for declining cigarettes), and how to substitute new habits for the old ones associated with smoking (e.g., do not carry matches, keep change ordinarily used for buying cigarettes in a different place, carry gum, take deep breaths or pace about when the urge to smoke comes). Much of the information in the pamphlet was familiar, and it did not promise an instant cure without effort. However, by using numerous examples of techniques for coping with situations that are part of a student's daily life, it was hoped that the pamphlet would lead to mental rehearsal of the steps which facilitate quitting or decreasing smoking. The use of an antismoking drug, such as Lobeline, was also mentioned.

---

[2] While viewing the film, photographs of the experimental subjects were taken with a flash camera at approximately 3-minute intervals. It was hoped that an analysis of facial expressions would provide additional evidence of emotional arousal. However, the photographs turned out to be rather poor technically, and these data will not be reported.

[3] The moderate-fear film was excerpted from *Smoking and You* distributed by Contemporary Films, Incorporated. The operation sequence was taken from *One in 20,000* made for the American Temperance Society. It was used as part of the high-fear stimulus in studies by Niles (1964), Leventhal and Niles (1964), and Leventhal and Watts (1966).

[4] The pamphlet was based on a booklet currently being used by the Narcotics Division, New York City, in its antismoking clinics. When questioned in a follow-up interview, only one subject said he had used an antismoking drug. Hence the effect of the drug recommendation on smoking behavior may be ignored.

Dependent Measures

Postquestionnaires. Immediately after the communication subjects filled out a third questionnaire. This questionnaire contained the same list of mood adjectives and the same items on attitudes toward smoking presented in the preexposure questionnaires. Two items on the fairness and accuracy of the film and the effectiveness of the recommendations were also included. A fourth questionnaire was administered to subjects individually 1 week after the communications. This questionnaire contained many of the same attitude items presented in the other two attitude questionnaires as well as items dealing with subjects' retrospective evaluation of the program and with changes in subjects' smoking behavior.

Behavior checks. Five estimates of the average number of cigarettes smoked per day were obtained from subjects. For each estimate subjects reported their average daily consumption over a period of 1 week, except for the "vacation" measure, described below, for which the period was 2 weeks. The first report referred to the week immediately preceding the program and was made on the precommunication questionnaire. The second report was made on the follow-up questionnaire and referred to the week immediately after the program. The third and fourth reports were obtained by telephoning each subject and asking how much he had smoked during the current week and over the Christmas recess which had preceded this week. The final report was made on a postcard sent to the subjects 3 months after the program.

RESULTS

There were no significant differences among conditions on the preexposure measures of mood, attitudes, or smoking habits. Difference scores were used in analyses of the attitude items and the data analyzed by means of analysis of variance using the unweighted-means formula for unequal *n*'s given by Winer (1962, pp. 241–244). The *n* for each analysis was the number of subjects who answered the item on both the pre- and postquestionnaire. Differences between the treatment groups and the control condition were assessed by *t* tests using the error term from the analysis of variance.

Emotion and Beliefs

Negative emotion. Since the measures of fear, depression, nausea, and shame were found to be highly correlated, all items related to these mood dimensions were summed to provide an overall index of negative emotion. Table 1 shows that the fear manipulations increased verbal reports of negative emotions as intended. Subjects exposed to the moderate-fear film reported significantly more negative emotion than control subjects who saw no film ($t = 3.14$, $df = 101$, $p < .005$), and subjects who saw the high-fear film reported more negative emotion than subjects who saw the moderate-fear film ($F = 17.03$, $df = 1/101$, $p < .005$). There were no main effects or interactions for the other manipulated variables or for the smoking levels.

Aggression. An index of aggression was derived by summing the scores for the mood items "sceptical," "suspicious," "defiant," and "fed-up." There was one significant effect on this index: subjects who did *not* smoke during the program expressed more aggression than subjects instructed to smoke (smoking condition, $\bar{X} = 1.25$; no-smoking condition, $\bar{X} = .86$; $F = 5.46$, $df = 1/101$, $p < .05$).

Vulnerability. One item from the mood questionnaire and one from the attitude questionaires were used to measure perceived vulnerability. On the mood item ("I feel vulnerable") there were significant main effects for both the fear and the specific instructions manipulation. Subjects in the high-fear condition reported feeling more vulnerable than subjects in the moderate-fear condition (high fear, $\bar{X} = 1.65$; moderate fear, $\bar{X} = .46$; $F = 9.29$, $df = 1/101$, $p < .005$), and subjects who did not receive specific

TABLE 1

CHANGES IN EMOTIONS AND ATTITUDES[a]

Items	Control	Moderate-Fear film	High-Fear film	High vs. moderate	Moderate vs. control
Negative emotions	−6.72	6.92	17.79	<.005	<.005
Aggressive feelings	−1.16	−.28	−.10	*ns*	*ns*
Vulnerability feelings	.22	.46	1.65	<.005	*ns*
Desire to quit smoking	.56	.39	1.13	<.005	*ns*
n	12	45	61		

Note.—The index of negative emotions is the sum of 14 difference scores. The index of aggressive feelings is the sum of four difference scores.

[a] All entries in the table are based on differences between precommunication and postcommunication scores and are unweighted averages of the means for the subgroups within each condition. Means for the film conditions are the average of eight group means. Means for the control condition are the average of two means (light and heavy smokers).

instructions on how to stop smoking reported feeling more vulnerable than subjects who received such instructions (instructions, $\bar{X} = .50$; no instructions, $\bar{X} = 1.41$; $F = 8.01$, $df = 1/101$, $p < .005$). On the attitude item ("How susceptible do you *now* feel to lung cancer?") there was a similar main effect for the instructions manipulation (instructions, $\bar{X} = .49$; no instructions, $\bar{X} = 1.02$; $F = 5.40$, $df = 1/101$, $p < .05$), no effect of the fear manipulation, and an interaction between level of smoking (light versus heavy smokers) and instructions to smoke during the program ($F = 7.39$, $df = 1/101$, $p < .05$). In this interaction heavy smokers reported feeling more susceptible to lung cancer when told to smoke during the program (smoking condition, $\bar{X} = 1.43$; no-smoking condition, $\bar{X} = .44$) than when not permitted to smoke, while there was no difference between conditions for light smokers (smoking condition, $\bar{X} = .46$; no-smoking condition, $\bar{X} = .71$).

Beliefs. Items measuring beliefs that smoking causes lung cancer ("The available evidence shows, without reasonable doubt, that smoking causes lung cancer"; "The less a person smokes, the less likely he is to develop lung cancer") and beliefs in the effectiveness of stopping smoking ("For a smoker, how effective is stopping smoking in reducing the chances of developing lung cancer?") were not significantly affected by the manipulations.

Desire to Comply with the Recommendation

Table 1 shows the mean differences between preexposure and postexposure scores (measured immediately after the communication) or intentions to decrease smoking. Subjects in the high-fear condition showed significantly larger increases in desire to quit smoking than subjects in the moderate-fear condition ($F = 7.41$, $df = 1/101$, $p < .05$), but there was no significant difference between the moderate-fear and control conditions.

Desires to quit smoking were also measured 1 week after the communications, and differences between these and the preexperimental scores were computed. The delayed measures show a pattern similar to that of the immediate measures, but the difference between the high- and moderate-fear conditions was no longer significant. Obviously, responses to the delayed items may have been affected by actual compliance during the week interval.

Credibility of Sponsor and Communication

All groups considered the Public Health Service (the alleged sponsor of the program) to be a highly reliable source of information on smoking and lung cancer. A more direct measure of message credibility was obtained from subjects in the film groups, who were asked whether they thought the movie gave a "fair and accurate" picture of smoking and lung cancer. Subjects instructed to smoke during the program were *more* favorable in their responses than subjects instructed not to smoke (smoking condition, $\bar{X} = 4.53$; no-smoking condition, $\bar{X} = 3.91$; $F = 4.23$, $df = 1/100$, $p < .05$). There were no other significant differences.

Smoking Behavior

The five reports of smoking behavior obtained from each subject were subjected to square-root transformation, and an analysis of variance for repeated measures was performed using the unweighted-means solution of Winer (1962, pp. 376–378). The analysis of variance involved the four experimental variables previously described and a fifth (stages) factor, corresponding to the five periods for which reports of smoking behavior were obtained. One or two of the five reports of smoking were missing for 10 subjects because the experimenter was unable to reach them at the appropriate times and these data were estimated. Four subjects who initially smoked fewer than two cigarettes were excluded from the analysis.

The experimental communications were effective in that the average number of cigarettes smoked per day decreased sharply over the follow-up period (stages—$F = 18.11$, $df = 4/388$, $p < .005$). The difference between the control and all experimental conditions taken together was highly significant ($p < .005$ for t tests) for each of the four reports of smoking made after the communications.

Fear Level

As can be seen from Figure 1, there was little difference in the reported smoking behavior of subjects in the high-fear and moderate-fear conditions at any period for which measures were taken after the communications (fear—$F = .25$, $df = 1/97$; Fear × Stages—$F = 1.12$, $df = 4/388$). The original differences in emotional arousal associated with the films seem to have had no influence on subsequent smoking behavior. However, some minimal amount of arousal seems necessary to motivate behavior change as shown by the significant difference between the moderate-fear and control conditions for each report of smoking following the communication ($p < .05$ for each of the four t tests).

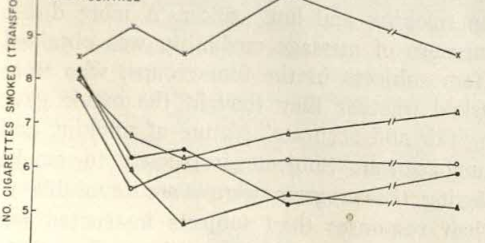


FIG. 1. Smoking behavior over time in fear and instructions conditions.

An interaction in the predicted direction between fear and instructions on how to quit smoking is also apparent in Figure 1 ($F = 3.50$, $df = 1/97$, $p < .10$). The receipt of instructions seemed to make more of a difference in the high-fear condition than in the moderate-fear condition. In the moderate-fear condition the instructions and no-instructions curves remain close together over the 3-month period, and subjects who received instructions reported less smoking than subjects who did not receive instructions only at the last two points in time. In the high-fear condition, on the other hand, subjects in the instructions condition always reported less smoking than subjects in the no-instructions condition, and the difference between conditions tended to *increase* over time.

Instructions on How to Stop Smoking

When the behavior of subjects who received instructions on how to stop smoking was compared to that of subjects who did not receive instructions, the main effect was not significant ($F = 2.63$, $df = 1/97$), but the interaction with stages was significant ($F = 2.69$, $df = 4/388$, $p < .05$). Figure 2 shows that the instructions and no-instructions curves both show sharp drops in smoking from before the program to 1 week after, and then begin to diverge. The divergence of the two curves was tested by applying linear weights to the four postcommunication measures for each condition and testing the interaction of the linear components with stages. This interaction was significant ($F = 5.17$, $df = 1/388$, $p < .025$). During the period from 1 week to 3 months after the program, subjects who received

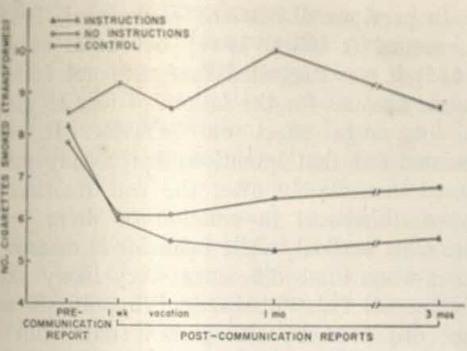


FIG. 2. Smoking behavior over time in instructions conditions.

instructions on how to stop smoking apparently maintained their decreased smoking level, while subjects who did not receive these instructions tended to regress toward their precommunication smoking level. For the differences between the instructions and no-instructions conditions t values were very small at 1 week ($t = .05$) and during vacation ($t = .74$), and larger at 1 month ($t = 1.67$, $df = 97$, $p < .10$) and at 3 months ($t = 1.67$, $df = 97$, $p < .10$).

Instructions to Smoke during the Program

Contrary to prediction, smoking during the film slightly decreased compliance with the recommendation to stop smoking. However, the main effect of this factor was not significant ($F = 2.51$, $df = 1/97$), nor was the interaction with stages ($F = .90$, $df = 4/388$).

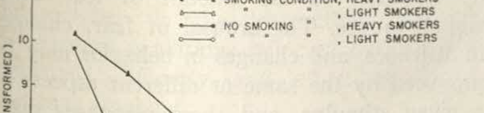


FIG. 3. Smoking behavior of heavy and light smokers in smoking conditions.

Light versus Heavy Smokers

The main effect of initial level of smoking was, of course, very large ($F = 104.57$, $df = 1/97$, $p < .005$) and simply indicates that heavy smokers continued to smoke more than light smokers after the program. Of more interest is the interaction of this variable with stages ($F = 3.07$, $df = 4/388$, $p < .05$). Light smokers showed little change in their smoking behavior after the 1-week stage, while heavy smokers reported a further decrease from 1 week to the vacation. Of interest also is the interaction between the light-heavy classification and instructions to smoke during the program ($F = 4.92$, $df = 1/97$, $p < .05$). Heavy smokers instructed to smoke during the film were less likely to decrease smoking afterward than heavy smokers instructed not to smoke during the film. It made little difference to light smokers whether or not they smoked during the film.

Supplementary Analyses

Reports of smoking behavior showed considerable change in variance over time. Some subjects never decreased smoking, others decreased at one or two stages but not at others, and still others decreased smoking at all four stages. The amount by which subjects decreased smoking also varied widely, with some subjects reporting they had quit completely and others reporting only token decreases. Supplementary analyses of smoking behavior were therefore undertaken using the proportions of subjects who reported substantial change in their smoking. This type of score gives a more stable picture of the impact of the communications on behavior since it ignores minor variations in an individual's smoking behavior and is unaffected by changes in within-group variances over time. The following procedure was used for these analyses: all light smokers who decreased smoking by at least three cigarettes (the median number of cigarettes cut down by light smokers in the total sample), and all heavy smokers who cut down by at least six cigarettes (the median number cut down by all heavy smokers) were counted as having truly decreased smoking. The proportions of decreasers in each condition for each stage were then computed and transformed to arc sines,[5] and a separate analysis of variance for each stage was carried out on the transformed proportions.

The results of these analyses proved to be very similar to those for the mean number of cigarettes smoked. For example, the difference between the proportions of decreasers in the instructions and non-instructions conditions was not significant at 1 week (instructions $= 51\%$, no instructions $= 40\%$; $F = 2.01$) or during the vacation (62% versus 34%; $F = 2.79$, $p < .10$), but was highly significant at 1 month (64% versus 34%; $F = 10.50$, $p < .005$) and at 3 months (53% versus 26%; $F = 9.51$, $p < .005$). The proportion of decreasers in the control condition was 9% at 1 week and 20% during the vacation, at 1 month, and at 3 months.

Discussion

Subjects saw one of two films on the dangers of smoking. The moderate-fear film presented evidence linking lung cancer to smoking and showed patients suffering from relatively mild disorders caused by smoking. The high-fear film differed from the moderate-fear film in including a section depicting a lung-cancer operation. The high-fear film increased reports of unpleasant emotions including fear, feelings of personal vulnerability to lung cancer, and desires to stop smoking, but it did not lead to a greater decrease in reported smoking behavior than the moderate-fear film. In turn, subjects exposed to the moderate-fear film reported increased fear and decreased smoking compared to control subjects who received only the recommendation to stop smoking and the instructions on how to stop smoking.

In previous discussions (Leventhal, 1965; Leventhal & Niles, 1964; Leventhal et al., 1965) it was suggested that temporal factors might account for the failure of fear to exert a long-range effect on behavior. It was pointed out that intentions are usually measured immediately after the fear treatments when differences in emotion or drive level are most marked, while behavior is measured later when these differences very likely have dissipated. This "dissipation" hypothesis was not disproven by the present data, but it seems too simple to account for the observed effects. The major finding which it fails to handle is that the difference in reported fear between the control and moderate-fear condition *was* associated with differences in reported behavioral compliance over time, while the difference in fear between the moderate-fear and high-fear condition was ***not*** associated with differences in compliance. Thus an increase in fear produced by relatively factual depiction of the dangers of smoking (the moderate-fear communication) had a long-range positive effect on compliance, but little effect on immediate intentions to comply, whereas a further increase in fear produced by the shock tactics of the high-fear film exerted an immediate effect on intentions to comply, but had no long-range effect on behavior.

This finding raises the question whether the inconsistent effects among studies noted previously may not be better understood by an analysis of informational and stylistic qualities of the various stimuli, rather than by reference to the amount of fear these stimuli arouse. The arousal of fear, changes in attitudes, and changes in behavior may be produced by the same or different aspects of a given stimulus, and these responses may, in turn, be merely correlated rather than causally related.

That the important next step in research is a more refined analysis of communications is also suggested by the effects on persuasion of a second experimental variable—the receipt of instructions on how to stop smoking. In this study, as in the previous tetanus studies (Leventhal et al., 1965, 1966), the receipt of instructions on how to cope with the danger apparently decreased feelings of

---

[5] Thanks are extended to Robert Abelson for suggesting the use of the arc-sine transformation. The error term for this test is:

$$\frac{1/nl + \ldots\ldots 1/nk.}{\text{number of cells}}$$

Comparisons between the film groups and the controls were performed using the formula:

$$Z = \phi_1 - \phi_2 / \sqrt{1N_1 + 1N_2}.$$

It should be noted that chi-square tests of the main effects gave p values similar to those obtained using the arc-sine transformation.

vulnerability to the danger, increased behavioral compliance with protective recommendations (so long as the instructions were accompanied by at least a moderately fear-arousing communication), and had no discernible effect on fear level or intentions to comply with the recommendations. Although these instructions seemed to aid high-fear subjects more than moderate-fear subjects, the effect of receiving instructions was apparently not due to differences in fear reduction. Possibly the instructions were effective because they increased the subject's confidence in being able to help himself, his awareness of the conditions which led him to smoke, and the techniques which could be used to inhibit smoking. By providing excuses for deferring cigarettes and preparing alternatives to smoking, the instructed subject could control his behavior well before he started smoking. The uninstructed subject, on the other hand, might just as frequently have remembered his resolution not to smoke, but he might have remembered on less appropriate occasions, for example, just after buying cigarettes. Thus the ego of the uninstructed subject must invent its own devices for avoiding smoking, and it will frequently be asked to control smoking when incentives for smoking are very powerful. The instructed ego does not need to invent evasive maneuvers to overcome strong temptation, and the result may be a more efficient execution of avoidance behaviors which readily become automatic response sequences unrelated to conscious decision making. It is possible that the postcommunication differences in reported vulnerability in the instructions and no-instructions conditions reflected differences in perceived difficulty and effort needed to decrease smoking.

Finally, a word should be said about the effects of smoking during the communication. It was expected that subjects who were instructed to smoke while the dangers of smoking were depicted would later show more compliance with the recommendation to decrease smoking. Quite the opposite proved to be the case. Subjects who smoked during the films reported somewhat less compliance with the recommendation than subjects who were not permitted to smoke. A possible explanation for this finding is that smoking during the film reminded subjects of the pleasures of smoking as suggested by the fact that subjects who smoked during the film reported less annoyance than subjects who did not. Another possibility is that smoking during the film reduced the credibility of the communication by reminding subjects that they had suffered none of the alleged consequences of smoking, or by suggesting that the communicator did not really regard smoking as dangerous. The present data do not lend support to either of these explanations, but perhaps the appropriate questions simply were not asked.

REFERENCES

JANIS, I. L., & TERWILLIGER, R. An experimental study of psychological resistances to fear-arousing communications. *Journal of Abnormal and Social Psychology*, 1962, **65**, 403–410.

LEVENTHAL, H. Fear communications in the acceptance of preventive health practices. *Bulletin of the New York Academy of Medicine*, 1965, **41**, 1144–1168.

LEVENTHAL, H., JONES, S., & TREMBLY, G. Sex differences in attitudes and behavior change under conditions of fear and specific instructions. *Journal of Experimental Social Psychology*, 1966, **2**, 387–399.

LEVENTHAL, H., & NILES, P. A field experiment on fear arousal with data on the validity of questionnaire measures. *Journal of Personality*, 1964, **32**, 459–479.

LEVENTHAL, H., SINGER, R. P., & JONES, S. Effects of fear and specificity of recommendations. *Journal of Personality and Social Psychology*, 1965, **2**, 20–29.

LEVENTHAL, H., & WATTS, J. C. Sources of resistance to fear-arousing communications. *Journal of Personality*, 1966, **34**, 155–175.

NILES, P. The relationship of susceptibility and anxiety to acceptance of fear-arousing communications. Unpublished doctoral dissertation, Yale University, 1964.

NOWLIS, V. Research with the Mood Adjective Check List. In S. Tomkins & C. Izzard (Eds.), *Affect, cognition, and personality*. New York: Springer, 1965. Pp. 352–389.

PAGANO, F. An experimental inquiry into the effectiveness of an educational approach in the prevention of smoking. Unpublished master's thesis, Yale University, 1965.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received August 17, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 322-327

BRIEF ARTICLES

RELATIVE INFLUENCE OF POSITIVE AND NEGATIVE INFORMATION IN IMPRESSION FORMATION AND PERSISTENCE

MARJORIE H. RICHEY, LUCILLE McCLELLAND, AND ALGIMANTAS M. SHIMKUNAS
Saint Louis University

University students were given inconsistent positive and negative blocks of written narrative information from which to rate the character of a stranger. ½ received the information in positive-negative order and ½ in negative-positive. Ratings were made after each block of information and again 7 to 9 days later. Initial ratings based on single univalent paragraphs were significantly altered in both groups by subsequent incompatible information, but the change was not equally permanent for both orders of presentation. Whereas originally positive impressions were lastingly changed by negative information, originally negative impressions which had been revised upward became significantly more negative again within 9 days. Replication with different information about the stranger (inversion of original content) yielded the same results.

In two frequently-cited studies by Luchins (1957a, 1957b) it was found that subjects given two opposing narrative descriptions of an unknown other formed impressions of this person which tended toward a univalent resolution of the inconsistency: one or the other description tended to dominate the impression. The order and manner in which the descriptions were given determined which was dominant; if both were given in immediate succession the first had more influence, but if a diverting task were interpolated between them the second predominated.

The descriptions used by Luchins (one describing the "other" as introversive and one describing him as extraversive) are objective behavioral descriptions which appear to have few evaluative or ethical connotations, and subjects were not specifically asked to evaluate the behavior. Rosencrantz and Crockett and their coworkers (Mayo & Crockett, 1964; Meltzer, Crockett, & Rosencrantz, 1966; Nidorf & Crockett, 1965; Rosencrantz & Crockett, 1965) have subsequently studied the resolution of inconsistent information which has evaluative implications for the other's character. They have been chiefly concerned with the effects of the variables of cognitive complexity (of subjects) and variations in the presentation of information upon the ability to form multivalent, as opposed to univalent, impressions. A secondary interest has been in the relationship of value congruence (between subject and other) to the favorability of the impression formed of the other.

In the present study, the problem investigated was whether, over and above differences in favorability produced by subject and procedural variables, there is a difference in relative influence of the opposing communications themselves—that is, whether either positive or negative information has an intrinsically stronger "pull," even though the two communications may be equally extreme if considered separately.

In a study which is tangentially related, Pastore (1960) asked subjects to rate the possibility of change of a person who possessed liked traits into one who possessed disliked traits or the reverse. He found that the first possibility was considered more likely. While this finding is relevant, it does not answer the question as to whether one type of contemporary characteristic, positive or negative, has a greater influence on evaluations.

A second question of the present study had to do with procedure. Studies of propaganda methods have shown that the full effects of a communication are not always immediately apparent. For example, Hovland and Weiss (1951) found that some subjects who had initially been unmoved by information from an untrustworthy source later did change their opinions in accordance with this information, apparently dissociating it from its source. Impression-formation studies of the type cited have dealt only with measures taken at the time of initial exposure to the conflicting information. The degree to which such impressions persist has not been investigated.

To summarize, the present study specifically concerned (*a*) the valence of impressions formed from inconsistent positive and negative information about another's character when systematic effort has been made to control extremeness as well as order of the opposing descriptions, and (*b*) the persistence of initial impressions through time.

METHOD

Subjects

Two experiments were performed, of which Experiment II was a replication of Experiment I, using a different form of the test instrument as described below. In Experiment I, subjects were 32 university students, 18 males and 14 females, who were enrolled in an undergraduate course in social psychology or a graduate course on exceptional children. The age range was from 20 to 35 years, with a median of 21. In Experiment II, subjects were 32 members of a general psychology class, comprising 22 females and 10 males. The age range was 18 to 50 years, with a median of 22. No attempt was made to equate the treatment groups in either experiment for age or sex.

Instrument

The instrument was an original one constructed for this study and suggested by that used by Luchins (1957b) with introversion-extraversion data. The information given the subjects about the other, or person to be judged, in this study consisted of five positive and five negative statements relating particular interactions with business colleagues and family members. Two content forms were used, Form I in Experiment I and Form II in Experiment II. Each involved the same general situations, but in Form I the other behaved positively (ethically or unselfishly) in the same situations in which he behaved negatively (unethically or selfishly) in Form II, and vice versa. Form II was developed and Experiment II performed to determine whether or not the results of Experiment I would be obtained from different specific content. The two content forms are as follows:

Form I

Positive I. C risked his job by protesting against an unfair personnel practice of his company; all employees benefited from the resulting decision to change the practice. If he has not had time to give his secretary advance notice, he works overtime himself rather than imposing on her to stay after hours to complete rush assignments. When he saw that his wife was working too hard, he quietly sold a cherished stamp collection in order to hire help for her, telling her that he had lost interest in the collection. He took a demotion in rank (his salary remained the same) in order to move to his present location where there are superior treatment facilities for a health problem of his daughter, although she would probably have been able to get along fairly adequately in the other location. He writes and telephones his elderly parents frequently and goes to stay with ther when either is seriously ill.

Negative I. When C wanted to get rid of a subordinate whom he perceived as a potential rival, he did not give the subordinate a merited salary increase, thereby eventually causing the subordinate to resign. He sometimes utilizes the ideas of more creative colleagues and subordinates without giving them credit, acting as though he were the originator. At a party which he and his wife attend together, he tends to ignore her and flirt with any younger women who happen to be present. He has little time for activities with his children; when at home he is "too tired." He contributes nothing to the financial support of his elderly parents, who live on a meager pension, although he buys a new car for himself every year.

Form II

Positive II. Although C would have been happy to be rid of a subordinate whom he perceived as a potential rival, he gave good reports of this man's work to their superiors and gave him a deserved salary increase without which the subordinate eventually would have quit. He never accepts credit for ideas of more creative colleagues and subordinates in his department but always gives full public recognition to the originator, even when he could pretend the ideas were his without the originator's knowing it. At a party which he and his wife attend together, he appears to enjoy her company and passes up the opportunity to flirt with younger women who find him interesting. He makes time for activities with his children even when he is tired and pressured by other demands. He contributes to the financial support of his elderly parents, who would otherwise have to live on a meager pension, although it means that he must forego some luxuries enjoyed by others in comparable positions.

Negative II. C supported an unfair personnel practice of his company because he thought protesting might hurt his position with his superiors; all employees were penalized by continuation of the practice. Whether or not he has given his secretary advance notice, he frequently has her stay after hours to complete rush assignments, although he does not work overtime himself. When he saw that his wife was working too hard, he might have obtained money to hire help for her by selling a cherished stamp collection, but instead he kept it and continued purchasing more stamps to add to it. He would not accept a transfer (his salary would have remained the same) in order to move to another location where there are superior facilities for treatment of a health problem of his daughter, who has usually managed so far to get along where she is. He rarely writes or telephones his elderly parents and did not go to visit them during a recent serious illness of his mother.

In pilot studies, the attempt had been made to develop positive and negative paragraphs of equal extremeness or intensity. Pilot subjects were given a single paragraph (either P or N) from which they rated their impression of the other's character on a 7-point scale. Revisions were made in the materials until mean scale values of approximately

2 were obtained for the negative paragraphs and approximately 6 for the positive ones. The resulting paragraphs, therefore, when rated singly by pilot subjects, were of values approximately equidistant from a scale value of 4, defined as "average; neither a better nor worse impression than I have of the average person." The Form II paragraphs are not exact inversions of Form I. The experimental design required that the Form II paragraphs be rated equally as extreme as the Form I. In some cases, the other's not committing a particular negative action yielded only an average or mildly positive rating; to get a positive rating of equal extremeness, the other had to take some additional positive action as well as not committing the negative one. The reverse was also true; his not taking a particular positive action did not always cause subjects to censure the other. The two forms, therefore, while referring to similar situations, necessarily involve some different specific content elements.

Procedure

Subjects in each experiment were assigned randomly to one of two orders of presentation, positive-negative (PN) or negative-positive (NP). Subjects in both order groups were given a first block of information (either P or N) and then asked to rate their impression of the other's character on the 7-point scale described above. When they had completed this initial rating, all subjects were given a task intended as a diversion. The task assigned was the drawing of two human figures, one male and one female. Fifteen minutes were allowed for the drawings. Following this interval, each group was given the second, incompatible block of information about the other. The task was introduced with the sentence: "The following is additional information concerning the same person about whom you read on page 1." After they had read the second paragraph, they again rated the other, this time under instructions to "use all the information now available to you." Immediately after completing the ratings, they were asked to write a paragraph describing their impression of the other's character.

In order to assess the permanence of the impression, 7 to 9 days later all subjects were requested (without prior notice) to rate the other a third time. They were also asked to list everything they could remember about him.

RESULTS

The mean deviations from 4 (midpoint of the rating scale) of the initial ratings, based on single blocks of information, were compared for each experiment. Results indicated no difference in extremeness between P and N paragraphs in either experiment ($t = .33$, $p = > .10$ in Experiment I; $t = .70$ $p = > .10$ in Experiment II).

Data from the two experiments were then analyzed simultaneously in a $2 \times 2 \times 3$ trend analysis with two levels of content (form of test instrument), two orders of presentation, and three trials. Results are presented in Table 1. As the table indicates, there was no differential effect of content. There was a main effect of order, a main effect of trial, and a significant interaction of trial and order.

TABLE 1

ANALYSIS OF VARIANCE OF CHARACTER RATINGS

Source	df	MS	F
Content (A)	1	.02	.01
Order (B)	1	8.33	4.27*
A × B	1	2.52	1.29
Error (a)	60	1.95	
Trial (C)	2	7.65	6.49**
A × C	2	.90	.76
B × C	2	183.77	156.03**
A × B × C	2	.02	.02
Error (b)	120	1.18	

* $p < .05$.
** $p < .01$.

Table 2 gives the mean ratings for each treatment group for each trial. When the means were compared, using Duncan's multiple-range test and $\alpha = .05$, the order groups showed the expected difference in the P and N paragraphs considered singly, plus a significant recency effect on the second trial (i.e., greater influence of the more recent information), and no difference on the third ($PN_1 > NP_1$, $NP_2 > PN_2$, $NP_3 = PN_3$). Changes over trials are summarized as follows: in all groups, there was a significant change in impression after reading the second, incompatible block of information. The negative opinions changed toward the positive and vice versa ($NP_2 > NP_1$, $PN_1 > PN_2$). During the week that followed, however, the two NP groups *again* made significant changes, this time reverting toward their original negative opinions ($NP_2 > NP_3$). The PN groups also showed a tendency to change back in the direc-

TABLE 2

MEANS AND STANDARD DEVIATIONS OF CHARACTER RATINGS

Group	Trial					
	1		2		3	
	M	SD	M	SD	M	SD
Experiment I						
NP	1.88	1.36	4.81	1.64	3.31	1.30
PN	6.25	1.00	2.69	1.40	3.00	1.59
Experiment II						
NP	1.88	.81	4.94	1.23	3.81	.83
PN	5.88	1.02	2.31	.60	3.00	1.10

tion of their original ratings, but neither of these changes was significant ($PN_2 = PN_3$).

The paragraphs written by the subjects to describe their impressions of the other were sorted according to whether and how the subjects had resolved the inconsistent information into some sort of unified, organized impression. Seven a priori categories were set up which were intended to exhaust the logical possibilities of such a resolution, as follows:

1. Character of other basically good, with some negative qualities.
2. Character of other basically poor, with some positive qualities.
3. Character basically good, but sometimes tries to appear worse than he is.
4. Character basically poor, but sometimes tries to appear better than he is.
5. Character basically conflicted, neurotic, confused, unstable.
6. Character essentially typical or average, with some good and some poor qualities.
7. Others. For example, statements that it was impossible to reconcile the two descriptions; interpretations that did not deal directly with internal inconsistencies in the material, such as statements that the other had undergone a change in character.

In Table 3 are presented the results of the sorting of the descriptive paragraphs by one of the investigators. It can be seen that the modal explanations are different for the two experiments. Subjects in Experiment I most often gave Explanation Number 4, stressing insincerity. In Experiment II, on the other hand, the modal explanation was Number 6—"essentially typical or average, with some good and some poor qualities." It should be recalled that these paragraphs were written in the initial testing session and before the "sleeper" effect had appeared in the quantitative data.

TABLE 3

FREQUENCIES WITH WHICH DIFFERENT EXPLANATIONS OF OTHER'S CHARACTER WERE OFFERED BY SUBJECTS IN BOTH EXPERIMENTS

Explanation	No. Subjects Offering					
	Experiment I			Experiment II		
	PN	NP	Total	PN	NP	Total
1. Basically good	1		1			
2. Basically poor				3		3
3. Tries to appear worse						
4. Tries to appear better	5	7	12	4	2	6
5. Conflicted, unstable	4	4	8	3		3
6. Typical, average	2	2	4	2	10	12
7. Other	4	3	7	4	4	8
Totals	16	16	32	16	16	32

At the time the follow-up ratings were made, subjects had also been asked to list everything they could remember about the other. A count was made of the positive and negative memories so recorded on the assumption that, since the final ratings had been predominantly negative, the negative behaviors of the other might have been better remembered. For each experiment, the numbers of subjects having more negative than positive memories were compared with those having more positive than negative. When a sign test was applied to these differences, they were found to be insignificant in both cases ($N = 21$, $X = 8$, $p = .38$ and $N = 21$, $X = 6$, $p = .08$ in Experiments I and II, respectively).

DISCUSSION

The findings are interpreted as evidence for the salience of negative information when negative and positive are presented in equal amounts as the basis for forming impressions of character.

While both initially positive and initially negative impressions were significantly modified by subsequent inconsistent information, the change was more permanent when it was in the negative direction. With the passage of time (within the first 9 days) the negative impression which had been revised upward underwent a significant reversion toward its original negative value, but the positive impression revised downward was not restored to a significant degree in the same period.

Since the content forms elicited different modal explanations of the other's character while showing the same downward trend in evaluation, it appears probable that it is not a specific negative content, but rather negative information per se which has this effect. At least two explanations may be suggested. From the qualitative data, especially in Experiment I, it would appear that negative behavior is viewed as intrinsically more genuine or sincere than positive. The other was often accused of being "phony" or "putting on a good front" to impress others, but never of trying to appear worse than he was.

It is also possible that a person who behaves unethically as often as he behaves ethically is statistically "worse than average." Perhaps one (correctly) expects other people to behave in a positive or neutral manner most of the time, with only occasional lapses. The other who is half the time commendable and half the time reprehensible may indeed be somewhat atypical. To some subjects such behavior was evidence of emotional maladjustment, but such marked unpredictability or undependability, for whatever reason it occurs, may be enough to make a poor impression.

The explanations of negative salience suggested above should, however, already apply at the time of initial exposure to the conflicting data. They do not explain why there should be a greater negative than positive delayed or "sleeper" effect. This effect was apparently not related to better memory for negative items, since there was no difference in numbers of subjects recalling more positive or more negative items. Possibly it simply takes more time than a few minutes to weigh such considerations as probable sincerity adequately. Another possibility, however, is that subjects have a wary set in character evaluation which is somewhat suspended in the immediate presence of positive information but reasserted whenever such information is later recalled in a larger context. To the extent that this is true, subjects may have deliberated very little about the other, either at the time they made their final ratings or during the preceding week, but rather may have responded more intuitively than logically.

In recent papers (Fishbein, 1963; Fishbein & Hunter, 1944), Fishbein has contrasted two different theories of the relationship between knowledge about an object and attitude toward that object. One of these, which he calls the averaging or consistency theory, holds that the value of the attitude toward an object is a function of the mean amount of affect contributed by all an individual's beliefs about the object. The other theory proposes that the affect contributed by the various beliefs summates, so that an attitude becomes increasingly intense as additional information of similar valence is acquired. As Fishbein and Hunter (1964) indicated, both theories would make similar predictions regarding the resolution of equal amounts of positive and negative information. Neither theory, as interpreted by the present writers, would predict the present finding of a greater weight of negative data in combination with positive.

The explanations offered for this finding would seem to fit better some type of contextual or "change in meaning" hypothesis (cf. Asch, 1946). Behavior which by itself was seen as ethical and unselfish was interpreted by some subjects as evidence of insincerity or emotional maladjustment when viewed in the context of an equal amount of negative behavior. The finding, moreover, suggests that context may affect the meaning of some kinds of information more than others; that is, the meaning of positive information is changed more permanently by the addition of negative data than the meaning of negative information is changed by the addition of positive.

Although replication of the results of Experiment I with the different content of Experiment II provides some basis for generalization of the findings beyond the degree possible with only one content form, it is still possible that the results obtained are a function of the particular contents used. To what extent the findings are the result of using strongly positive and strongly negative statements about the other is another question which is not answered. Research is underway to investigate the effects of such stimulus characteristics as quantity and extremeness of opposing information as well as contextual variables such as the relationship between subject and other.

No attempt is made to generalize these findings to content without evaluative implications for character. It seems possible that with regard to judgments of some attributes, for example, intelligence and talent, positive or favorable interpretations of conflicting data would be more plausible than negative, inasmuch as it would seem less likely that one could "fake" these qualities than character.

REFERENCES

ASCH, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, **41**, 258–290.

FISHBEIN, M. An investigation of the relationships between beliefs about an object and the attitude toward that object. *Human Relations,* 1963, **16**, 233–239.

FISHBEIN, M., & HUNTER, R. Summation versus balance in attitude organization and change. *Journal of Abnormal and Social Psychology,* 1964, **69**, 505–510.

HOVLAND, C. I., & WEISS, W. The influence of source credibility on communication effectiveness. *Public Opinion Quarterly,* 1951, **15**, 635–650.

LUCHINS, A. S. Experimental attempts to minimize the impact of first impressions. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* New Haven, Conn.: Yale University Press, 1957. Pp. 62–75. (a)

LUCHINS, A. S. Primacy-recency in impression formation. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* New Haven, Conn.: Yale University Press, 1957. Pp. 33–61. (b)

MAYO, C. W., & CROCKETT, W. H. Cognitive complexity and primacy-recency effects in impression formation. *Journal of Abnormal and Social Psychology,* 1964, **68**, 335–338.

MELTZER, B., CROCKETT, W. H., & ROSENCRANTZ, P. S. Cognitive complexity, value congruity, and the integration of potentially incompatible information in impressions of others. *Journal of Personality and Social Psychology*, 1966, **4**, 338-342.

NIDORF, L. J., & CROCKETT, W. H. Cognitive complexity and the integration of conflicting information in written impressions. *Journal of Social Psychology*, 1965, **66**, 165-169.

PASTORE, N. A note on changing toward liked and disliked persons. *Journal of Social Psychology*, 1960, **52**, 173-175.

ROSENCRANTZ, P. S., & CROCKETT, W. H. Some factors influencing the assimilation of disparate information in impression formation. *Journal of Personality and Social Psychology*, 1965, **2**, 397-402.

(Received April 27, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 327-331

FEAR REDUCTION FOLLOWING OBSERVATION OF A MODEL[1]

JAMES H. GEER AND ALAN TURTELTAUB

University of Pennsylvania

60 female undergraduates, classified as having high fear or low fear of snakes, were asked to approach a snake twice. Between approaches 20 Ss observed, in the same situation, a confederate who acted quite frightened. 20 Ss observed a confederate who acted calmly, and 20 Ss simply waited 5 min. between approaches. Using the change in distance from the snake as the dependent variable, high-fear Ss approached closer following observation of a calmly acting confederate. The results are discussed in terms of a social learning theory.

Fear or anxiety plays a central role in many theories that deal with human behavior, and the present study was designed to further knowledge concerning fear. Fear, for the purpose of this study, is conceived of as a negative emotional response to a specific stimulus. The study is one of a group (cf. Geer, 1965, 1966) that has attempted to investigate fear using objective behavioral measures. The fear-arousing situation used in most of the studies from this laboratory is analogous to animal investigations where variations in approach to a fear stimulus were taken as the measure of fear. A similar situation has been used as a criterion measure in some investigations of behavior therapy (cf. Lang, Lazovik, & Reynolds, 1965). Geer (1965), using approach to a fear stimulus as the dependent variable, reported that the sex of the experimenter and a time delay prior to approaching a fear stimulus were not effective in modifying fear behavior. An unpublished study by the same investigator reveals that different amounts of prior exposure to the fear stimulus do not affect subsequent approach behavior. The present study, also using approach to a fear object, reports the manipulation of fearful behavior where the independent variable was a social stimulus.

The effect of the presence of another person upon emotional behavior has been suggested by numerous investigators. Jones (1924) reported that the presence of a child who showed no fear of an object that was frightening to another child reduced the fear of the fearful child. She called this phenomenon fear reduction by social imitation. Freud (1955), using clinical evidence, noted that fear reduction sometimes occurred when phobic individuals were accompanied by "specially selected persons [p. 115]." Schachter (1959) studied affiliative behavior under conditions of anxiety arousal. In general, he found that subjects under conditions of anxiety arousal tend to affiliate. Schachter (1959) said, "People do serve an anxiety-reducing function for one another [p. 26]." Ring, Lipinski, and Braginsky (1965) reported a study in which the subject's moods were measured under conditions where confederates accompanying the subject behaved, depending upon experimental conditions, "extremely anxious," "moderately anxious," "somewhat anxious," or "low in anxiety." Ring et al. reported that the emotional state of the confederate has differential effects upon the subject's responses to mood questionnaires, and that this effect varies with the subject's birth order.

In the present study an attempt was made to assess the effect of observing the behavior of others in a fear-arousing situation upon the subject's subsequent reaction in that situation. The

---

[1] The data for study were collected by the junior author as part of an undergraduate research project. The authors wish to thank Janice Mazer and Jeanne Salis who acted as confederates in this study.

hypothesis under examination was that observation by a fearful subject of either fearful or nonfearful behavior by a model would modify the observer's behavior in a direction to make the subject's behavior more like the model's behavior.

METHOD

Subjects. Subjects were 64 female undergraduates enrolled in introductory psychology at the University of Pennsylvania. About 1 month prior to the experiment, students in the course had been given the Fear Survey, Schedule-II (Geer, 1965). Class members were classified as high fear if their response to the item, "snakes," was either "terror" or "very much," and low fear if their response to the item was either "none" or "very little." From the pool of high-fear and low-fear students, 30 low fear and 35 high fear were selected for participation in the study. Four high-fear subjects were discarded from the study because they failed to meet a behavioral criterion; that is, when they were first asked to approach a snake they touched it. One high-fear subject could not be used because she refused to take the behavior test. The fear inventory was used to identify individuals who were likely to be, depending upon their response to the item, "snakes," either high or low fear; however, the criterion for high or low fear was the behavioral test.

Procedure. All subjects were met by a male undergraduate who acted as the experimenter in the study, and were told that they were to participate in a study that had two purposes. One purpose was to study people's reactions to a snake upon two occasions. The second purpose of the study was described as an evaluation of the accuracy of judging emotional reactions to a situation as a result of the rater's having had experience in that same situation. Subjects were told that they would be rating the reactions of another subject who was approaching the snake. Subjects were further told that their judgments would be compared with ratings made by observers who had not had the same experience as the individuals being rated. There were, however, no other raters for such a comparison.

Following this introduction, subjects were told that at the far end of the next room on the floor was a live nonpoisonous snake.[2] The subject was asked to accompany the experimenter into the room and to stand near the wall of the room opposite the snake. After the subject was properly located (approximately 17 feet from the snake), the experimenter walked up to the small 6-inch wire barrier behind which the snake was placed, stepped over the barrier, and bent down and touched the snake so that it moved. This procedure was employed to demonstrate that the snake was alive and not dangerous. After this demonstration, the experimenter moved back across the barrier and stopped about 12 feet from the barrier. The subject was then instructed to "Approach the snake as close as you comfortably can. Cross the barrier and bend down and touch the snake if you can." The experimenter pointed out that the task was not to be taken as a challenge in which the subject could show how much she could take, but she was to do what was comfortable. The subject began her approach on signal from the experimenter. If the subject did not touch the snake, the experimenter recorded the distance that she stopped from the snake. This was done by the experimenter noting the spot where the subject stopped. The actual distance was measured without the subject's knowledge, after the subject had left the room. The experimenter, using a stopwatch, also recorded the time from the start signal until the subject either touched the snake or was at the point of nearest proximity to the snake.

After the first behavior test, the subject was taken outside the room that contained the snake and conducted to a small room that had a one-way vision mirror looking into the room where the behavior tests were conducted. From this point the subject could see into the behavior test room; a stereo system was hooked up so that the subject could clearly hear all that occurred in the test room. The experimenter told the subject that he would get the next subject and would go through exactly the same procedure with the next subject that the subject had just experienced. The subject was then given a rating sheet that she was to use in judging the reaction of the next subject. The items were:

1. Upon first viewing the snake, subject experienced the following amount of fear:
   None Very Little A Little Some
   Much Very Much Terror
2. How much tension or anxiety do you think subject felt when she was nearest the snake?
   None Very Little A Little Some
   Much Very Much Terror
3. Rate your impression of subject's overall fear level.
   None Very Little A Little Some
   Much Very Much Terror

The subject was also given a copy of Zuckerman's (1960) Affect Adjective Check List and instructed to check those items that described the emotional responses of the next subject. The subject was told to fill out the forms before the experimenter conducted an interview with the next subject. This interview, the subject was told, would be the criterion against which the subject's ratings would be compared. The rating sheet had two functions: (*a*) to make certain that the subject watched the next subject and (*b*) to have some means of assessing the effectiveness of the confederate. The subject was told that the one-way mirror was being used so that the next subject would not be aware of being observed. The subject was also told that during her second behavior test the observation room would be empty. The physical arrangement of the rooms was such that the subject could be fairly certain that she was not being observed during her second behavior test.

---

[2] The snake was a 4-foot pine snake loaned to the investigators by the Academy of Natural Sciences, Philadelphia, Pennsylvania.

The next subject was one of two confederates who behaved in one of two manners. One manner was termed high fear, and in that condition the confederate appeared quite frightened. The confederate stopped 15 feet from the snake during the approach, and during the interview the high-fear confederate spoke of considerable discomfort and asked several times to be allowed to leave the room. The low-fear confederate acted in a manner such as to indicate very little fear. It may be noted that most of the verbal descriptions of their feelings used by the confederates during the interview were taken from the responses of several pilot subjects who were interviewed in order to obtain responses for this purpose. The confederate did not know whether the observer was a high-fear or a low-fear subject, and the one-way mirror prevented the confederate from gaining any cues as to the subject's reaction. Confederates acted either the high-fear or low-fear role on the basis of a prearranged random schedule.

Following the observation, the experimenter went to the observation room, collected the rating sheet, and conducted the subject back for the second behavior test. The second behavior test was identical to the first, with the exception that the experimenter did not redemonstrate that the snake was alive. The same behavior measures were obtained in the second behavior test that were obtained from the first. In addition to the high-fear and low-fear confederate conditions, a third control condition was employed. This consisted simply of two exposures to the snake with a time lapse of 5 minutes (the mean time elapsed when confederates were used) between behavior tests. The subject sat in a separate room during this time. The subject sat alone without books or reading material. She was told that the delay was to provide two separate exposures to the snake. This condition was used to control for the effects of two successive exposures to the fear stimulus. Thus there were three groups of subjects—high-fear confederate, low-fear confederate, and control condition with both high-fear and low-fear subjects.

RESULTS

The first set of results pertain to the ratings made by the subjects when observing the confederates. The first measure was the summed response to the three items on the rating scale with 0 being assigned to the rating of "none" and the numbers 1, 2, 3, 4, 5, and 6 being assigned to ratings of increasing fear. The maximum possible score for the confederate was 18. The mean summed rating of the high-fear confederates was 12.9; the mean summed rating of the low-fear confederates was 3.1. The difference between those means was highly significant ($t = 10.47$, $p < .001$) indicating that the high-fear confederates were rated as showing more fear. A similar analysis was performed on the Adjective Check List. Using Zuckerman's (1960) scoring formula, the mean score for high-fear confederates was 15.1 and the mean score for low-fear confederates was 7.1. The difference between those means also yielded a highly significant difference ($t = 10.32$, $p < .001$). In sum, it may be inferred that the confederates were successful in transmitting the desired impression. One final analysis of the ratings was made. It would have been of interest if high-fear subjects attributed more fear to the confederate than did low-fear subjects. The data were consistently in the direction of suggesting that high-fear subjects did rate confederate as being more frightened. However, this "projection" did not approach accepted levels of statistical significance when evaluated by t tests.

The data of primary interest are the results of the distance measure. Table 1 contains a summary of the results for high-fear subjects. The table contains the median distances for high-fear subjects on Trials 1 and 2, and also the number of high-fear subjects who moved nearer the snake on Trial 2. The dependent variable used to assess the effect of the various conditions was each subject's change in distance from the first to the second behavior test. The measure was obtained only for high-fear subjects since all low-fear subjects touched the snake on both behavior tests. A Kruskal-Wallis nonparametric analysis of the three conditions indicated that the three conditions differed significantly from each other ($H = 6.64$, $p < .05$). Figure 1 presents the median-distance change scores for high-fear subjects under the three conditions. Mann-Whitney U tests between the groups revealed that the subjects moved closer to the snake following observation of the low-fear confederates than

TABLE 1

MEDIAN DISTANCE (INCHES) FROM SNAKE FOR HIGH-FEAR SUBJECTS AND NUMBER OF SUBJECTS THAT MOVED CLOSER ON TRIAL 2

	High-fear confederate	Control	Low-fear confederate	All high-fear subjects
Trial 1[a]	33	26.5	25	28
Trial 2	34	24.5	5	22.5
No. Ss moving closer on Trial 2	4	7	10	

[a] Mann-Whitney U test of differences between groups on Trial 1 did not approach statistical significance.

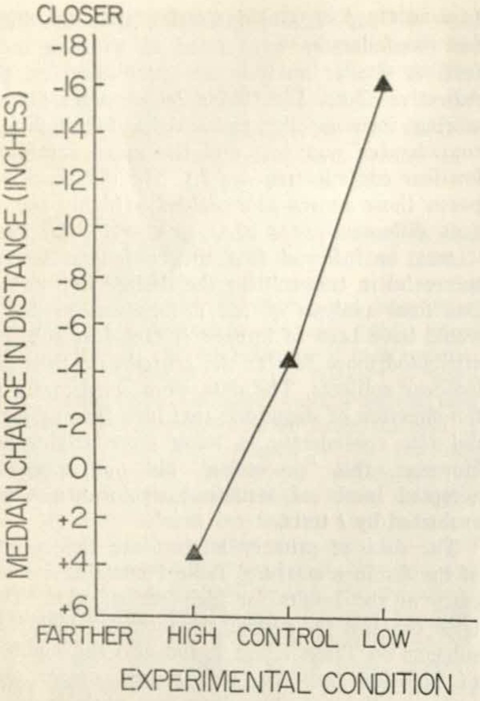


FIG. 1. The median change in distance for the high-fear subjects as a function of experimental condition.

after observing the high-fear confederates ($U = 22.5$, $p < .05$). The same test indicated that subjects moved closer to the snake following observation of the low-fear confederates than following the control condition ($U = 16.5$, $p < .02$). The same comparison between the control conditions and the high-fear confederate condition did not reveal a significant difference.

It may be recalled that the snake was behind a low barrier. It appeared that crossing that barrier was a difficult task for high-fear subjects since only 4 of the usable 30 high-fear subjects crossed the barrier on the first behavior test. Taking the results of high-fear control-condition subjects as the base line, it would be expected that on Behavior Test 2 one high-fear subject would cross the barrier who had not done so on Behavior Test 1. One additional crossing did occur in the high-fear confederate group, while six additional crossings occurred for the low-fear confederate group. A chi-square test of those data yielded a significant difference ($\chi^2 = 27.8$, $p < .01$) indicating that barrier crossings increased following observation of the low-fear confederate.

The trends of the latency data were all in the same direction as that of the distance data; however, there were no statistically significant differences. The reason for this appears to be in the nature of the task. Subjects were often noted to spend considerable time deciding whether or not to touch the snake, and this delay made the latency measure difficult to interpret and difficult to obtain accurately.

DISCUSSION

This study demonstrates that if an individual who displays fear of an object observes another person behaving unafraid of that object there is a tendency for the observer's fear behavior to be reduced. The obverse of this finding was not present in the data of this study; that is, subjects did not show significantly increased fearful behavior following observation of the high-fear confederate. Figure 1 reveals that there was a tendency for high-fear observers, having watched the high-fear confederate, to move further back on Behavior Test 2; however, the data did not attain accepted levels of statistical significance. This may be in part due to reverse psychology. Two high-fear subjects who observed the high-fear confederate reported that the other subject made them realize how "silly" their fear was, and thus on the second behavior test they showed much less fear than on the first test. It may also be that the subjects were quite frightened and only extreme measures would measurably increase the fear. Finally, it is possible that the confederates were unable to simulate fear accurately enough to affect the observer, while they could show low fear more effectively, since it was the confederates' normal response. Low-fear subjects could not move any closer to the snake upon viewing the low-fear model, since all low-fear subjects touched the snake on Trial 1. Viewing the high-fear model did not result in the low-fear subjects' moving away from the snake. Apparently the conditions of this study were not powerful enough to induce fear in low-fear subjects.

Several theoretical models may be employed to explain the results of this study. The most obvious to this writer is social learning theory such as discussed by Bandura and Walters (1963). They characterized one of the effects of a model's behavior upon an observer as a disinhibitory process. Disinhibition appears to be the major effect found in this study. Upon seeing a model behave in a nonfearful manner, the observer imitated the model and displayed behavior that had previously been inhibited, presumably because of fear. Social learning theory would further predict that the fearful models would have accentuated fearful behavior, however this did not occur. Several possible explanations of this failure were advanced above.

The present study does not indicate the mechanism by which behavior change occurs. A possible mechanism would involve self-instruction; that is, the observer, upon the occasion of the second behavior test, gave herself covert instructions such as, "If the other subject could do it, so can I." Indeed some subjects reported having such thoughts. It may be noted that such self-instruction could result in comfort responses which would function as competitors with (and thus reduce) fear. With fear thus reduced, approach behavior would continue. It is also possible that the subject's approach behavior was modified by self-instruction, but that other aspects of the fear response, such as autonomic components, were unmodified. Further studies are being conducted to investigate some of the possible mechanisms of the behavior change.

If it can be established that the changes in behavior that were reported in this study are relatively enduring, the results may have implications for the area of behavior modification. For example, the study suggests that fear reduction may be facilitated by the systematic and judicious use of models. Further, the effect of others upon fear may account for some of the results of group psychotherapy. This would seem to be particularly true if the members of the group had dissimilar fears.

REFERENCES

BANDURA, A., & WALTERS, H. *Social learning and personality development*. New York: Holt, Rinehart & Winston, 1963.

FREUD, S. *The standard edition of the complete psychological works of. . . .* (Trans. by J. Strachey) Toronto: Clarke Irwin, 1955.

GEER, J. H. Development of a scale to measure fear. *Behavior Research and Therapy,* 1965, **3**, 45–53.

GEER, J. H. Effect of fear arousal upon task performance and verbal behavior. *Journal of Abnormal Psychology,* 1966, **71**, 119–123.

JONES, M. C. The elimination of children's fears. *Journal of Experimental Psychology,* 1924, **7**, 382–390.

LANG, J., LAZOVIK, A. D., & REYNOLDS, D. J. Desensitization, suggestibility, and pseudotherapy. *Journal of Abnormal Psychology,* 1965, **70**, 395–402.

RING, K., LIPINSKY, C. E., & BRAGINSKY, D. The relationship of birth order to self-evaluation, anxiety reduction, and susceptibility to emotional contagion. *Psychological Monographs,* 1965, **79**, 10, 603.

SCHACHTER, S. *The Psychology of Affiliation*. Stanford, Calif.: Stanford University Press, 1959.

ZUCKERMAN, M. The development of an affect adjective check list for the measurement of anxiety. *Journal of Consulting Psychology,* 1960, **24**, 457–462.

(Received June 22, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 332-336

COGNITIVE CHALLENGE AS A FACTOR IN CHILDREN'S HUMOR APPRECIATION[1]

EDWARD ZIGLER
Yale University

JACOB LEVINE
West Haven Veterans Administration Hospital
Connecticut

AND LAURENCE GOULD
Yale University

The hypothesis was advanced that an important ingredient in the humor response is the degree to which the humor stimulus makes a cognitive demand on the individual. The specific prediction tested was that cartoons at the upper limit of the child's ability to comprehend would evoke the greatest amount of laughter as well as the greatest preference. Groups of cartoons varying in difficulty level from easy to nearly impossible were administered to 20 children (10 boys and 10 girls) of average intellect in each of the 3rd, 5th and 7th grades. As predicted, at all 3 grades mirth and preference scores peaked at the moderately difficult range of the difficulty dimension. The findings were discussed in relation to the child's motive of cognitive mastery, as well as to a motivation-free information-processing approach.

Despite the concern and investigatory efforts of philosophers and psychologists through the ages, no satisfactory comprehensive theory of humor has been forthcoming (see Flugel, 1954, and Levine, in press, for comprehensive historical reviews). Perhaps one reason for this failure is the frequently made assumption that humor is the result of some specific single factor. A more reasonable approach would appear to be one involving the view that humor is a phenomenon reflecting a number of causal factors. This notation of the complexity of the problem surrounding the humor response does not in itself carry us very far towards the problem's solution. What is required is the delineation and investigation of individual variables found to be related to humor, which should then be followed by the study of the interrelationships among such factors.

Some progress along these lines has been made in the body of work that has related the individual's psychodynamics to his tendency to laugh (Levine, in press). This work, for the most part, has emphasized the importance of the content of the humor stimulus, and how such content interacts with the individual's psychodynamics in determining the humor response. Although valuable, this approach has tended to underestimate the importance of cognitive factors in determining the degree of laughter. Whatever the individual's psychodynamic features may be, understanding a joke invariably requires the cognitive capacity to meet the intellectual demands posed by the joke. As obvious as this statement might be, the relation between intellectual level and response to humor stimuli has been far from clear (this literature reviewed in Zigler, Levine & Gould, 1966a).

This relation was investigated recently by Zigler et al. in a study with second-, third-, fourth-, and fifth-grade children. For children in grades two through four Zigler et al. were able to demonstrate that as the children's ability to comprehend a group of 25 cartoons increased, the magnitude of their mirth responses increased. However, contrary to prediction, although comprehension scores continued to go up, mirth scores for fifth graders showed a marked decline from that found for fourth graders. This study indicated that while comprehension is an important factor in determining the mirth response, perfect comprehension does not in and of itself guarantee laughter.

No evidence was found that the lowered mirth response of the fifth graders was due to the particular content of the cartoons. One explanation for this lowered mirth response is that the cartoons were too easy for the fifth graders and represented no intellectual challenge for these older children. It would appear that the child

---

[1] This research was supported by Research Grant MH 06809 from the National Institute of Mental Health, United States Public Health Service, the Gunnar Dybwad Award of the National Association for Retarded Children, and the Veterans Administration. The authors wish to express their appreciation to Inge Winer for running subjects and for her assistance in various phases of this study.

laughs at those cartoons which make appropriate demands on his cognitive structure, and not at those which are either too easy or too difficult. This suggests that an important ingredient in humor is the degree to which the humor stimulus makes a cognitive demand on the individual. Zigler et al. did find some evidence that cartoons which were at the upper limit of the child's ability to comprehend evoked the greatest amount of laughter. This generates the hypothesis that the mirth response is greatest when the complexity of the humor stimulus and the complexity of the child's cognitive structure are congruent.

Both Freud (1960) and Levine (in press) have also suggested that the sheer pleasure of employing ones' cognitive processes in relation to humor stimuli contributes to the experienced gratification. Inherent in this suggestion is the view advanced above that there is a factor influencing the magnitude of the mirth response that is independent of both the psychodynamic features of the individual and the individual's sheer cognitive ability to comprehend the joke. This factor may be related to what White (1959) has referred to as the effectance motive. Employing the effectance concept, one would expect the most laughter when the individual is experiencing the greatest sense of mastery, that is, at the point at which the cognitive demands of the humor stimuli are at the upper limits of the individual's cognitive ability (cognitively congruent). As noted above, some support for this view was obtained by Zigler et al. (1966a). However, their findings were somewhat equivocal since children at all grades received the same 25 cartoons.

A more direct test of the cognitive-congruence or effectance hypothesis suggests itself. Such a test requires that the difficulty level of the humor stimuli be equated across age groups. One could independently assess the difficulty of cartoons for each age level employed, and then categorize them as easy, moderately difficult, difficult, and nearly impossible. Given this procedure, the particular cartoons in each category would differ for each age level. The congitive-congruence hypothesis would generate the expectation that the mirth response should peak somewhere in the moderately difficult–difficult range at all age levels. In the present study this prediction was tested with third-, fifth-, and seventh-grade children of both sexes.

METHOD

Selection of Cartoons

This study required groups of cartoons whose difficulty level had been independently assessed on the basis of the comprehension scores of third-, fifth-, and seventh-grade children. The difficulty level of cartoons utilized in an earlier study (Zigler et al., 1966a) with third- and fifth-grade children was employed to select the easy, moderately difficult, and difficult groups of cartoons for these two grade levels. Grouping cartoons into difficulty levels for the seventh graders, for whom previous norms were not available, was accomplished by initially administering a heterogeneous pool of 35 cartoons to 20 seventh graders, 10 boys and 10 girls. On the basis of the comprehension scores obtained, groups of easy, moderately difficult, difficult, and nearly impossible cartoons were selected to be employed with the seventh-grade sample. The nearly impossible group of cartoons employed with the seventh-grade sample were also the cartoons classified as nearly impossible in the cartoon sets given to the third and fifth graders. In this way, four groups of four cartoons varying in difficulty level were selected for each grade. In addition to varying the difficulty within a grade level, the following criteria were employed in making the final selections: (*a*) For each grade the difficulty level within each group of four cartoons would be relatively homogeneous; (*b*) the difficulty level of any given difficulty grouping of four cartoons would be comparable across grades; (*c*) there would be no marked differences in the subject matter of the cartoons across grades or across difficulty levels within a grade. This last restriction was included in order to minimize a confounding of subject matter and difficulty. The results of the selection then are three different sets of cartoons for children in the third, fifth, and seventh grade which are comparable in difficulty and content. Each set is composed of four groups of four cartoons each, which on the basis of the children's comprehension scores are categorized as easy, moderately difficult, difficult, and nearly impossible. The order of presentation of each set of cartoons was randomized with the restriction that neither of the first two cartoons would be nearly impossible ones, and that two nearly impossible cartoons would not appear consecutively. The resultant order of cartoons for each grade was used for all subjects in that grade.

Subjects

A female experimenter administered the cartoons individually to 10 boys and 10 girls each in third-, fifth-, and seventh-grade classes (the mean CAs were 8.7, 10.6, and 12.6, respectively) in predominantly middle-class public schools. Selection of subjects was made by classroom teachers who were asked to select children who represented the typical or modal child in the class. They were specifically asked not to pick the brightest children, the dullest children, or children with any type of emotional or physical problem.

Administration

The examiner sat across from the child at a table with 16 cartoons mounted individually on cardboard lying face down on the table. The following instructions were given:

> I have a bunch of pictures, and I want to see if you think they are funny or not. As I show you each picture, tell me whether or not you think it is funny.

The experimenter then turned the cartoons over one at a time in a constant order and recorded: (*a*) the spontaneous mirth response of the child, (*b*) whether or not the child judged it funny, and (*c*) any spontaneous comments of the child. Following this initial presentation of the 25 cartoons, the experimenter said: "Now we'll go through them again, and this time you tell me about the cartoon. What is the joke? What is supposed to be funny about it?" On the second presentation, the experimenter scored for comprehension.

Following this second run-through the subjects were asked to select the three cartoons that they liked best.

Scoring

Three scores were obtained: (*a*) a facial mirth score employing the categories: 0 = negative response (grimace, etc.), 1 = no response (blank face, etc.), 2 = inhibited to a half or slight smile, 3 = full smile, 4 = laugh; (*b*) a comprehensive score employing the categories: 0 = no comprehension, 1 = partial comprehension, 2 = full comprehension; (*c*) preference ranking of cartoons based on the children's selections.

The interjudge reliabilities for the total mirth and comprehension scores in the earlier study (Zigler et al., 1966a) were found to be .95 and .94, respectively.

The subjects' comprehension and mirth scores for each level of cartoon difficulty are presented in Table 1. A 3×2×4 (Grade × Sex × Difficulty Level) repeated-measures analysis of variance of the comprehension scores resulted in a significant difficulty effect ($F_{3/162} = 252.39$; $p < .001$). None of the other main or interaction effects approached statistical significance. Summing across grade levels, Tukey tests were run between comprehension scores in adjacent difficulty groupings. Each of the three comparisons was found to be significant at beyond the .001 level. These findings indicate that the a priori grouping of the cartoons into difficulty levels accomplished the intended goals.

A 3 × 2 × 4 repeated-measures analysis of variance of the mirth-response scores resulted in a significant difficulty level effect ($F_{3/162} = 24.86$; $p < .001$). As can be seen in Table 1, the mirth response peaks at the moderately difficult level and then drops at the two higher difficulty levels at all three grades. A general nonmonotonic-trend analysis indicated that this particular pattern is a highly significant one ($F_{1/54} = 96.20$; $p < .001$), and accounted for the bulk of the variance associated with the difficulty level main effect. Individual comparisons run between adjacent mean mirth-response scores revealed that the mean for the moderately difficult cartoons was higher than that for the easy cartoons ($p < .05$); the mean for the moderately difficult cartoons was higher than that for the difficult cartoons ($p < .01$), which in turn was higher than that for the impossible cartoons ($p < .001$). It should be noted that the easy cartoons were not as easy as the impossible cartoons were difficult. Subsequent investigations should employ an easier group of cartoons. However, even lacking this symmetry in the difficulty dimension, the findings were in keeping with the cognitive congruence hypothesis.

The number of times that cartoons at each difficulty level were selected by the children as being among the three most preferred are presented in Table 2. If the three cartoons selected by each of the 20 children at a grade level were in the same difficulty category, that category would receive a score of 60. Across grade levels, then, the maximum score would be 180. As can be seen in Table 2, the preference scores are quite parallel with the mirth-response scores, rising from the easy category to peak at the moderately difficult category, and then dropping at the two higher levels of difficulty. The lack of independence of these preference scores precludes a statistical test of significance. In order to perform such a test, each child's preference score was converted to a rank score in the following way: (*a*) If all three of the child's selections were in the same difficulty category, that category would receive a rank of 1. Each of the other three categories would receive a rank of 3; (*b*) if two of the child's selections were in one difficulty category, and a third selection was in a second category, these categories would receive ranks of 1 and 2, respectively. The remaining two categories would receive a rank of 3.5; and (*c*) if the child made one selection from

TABLE 1

MEAN COMPREHENSION AND MIRTH SCORES FOR EACH LEVEL OF CARTOON DIFFICULTY

Grade	N	Cartoon difficulty level							
		Easy		Moderately difficult		Difficult		Impossible	
		Comprehension	Mirth	Comprehension	Mirth	Comprehension	Mirth	Comprehension	Mirth
3rd	20	1.85	2.16	1.51	2.34	1.15	2.12	.18	1.69
5th	20	1.80	2.12	1.59	2.18	1.15	1.71	.26	1.69
7th	20	1.71	2.22	1.39	2.34	1.08	2.11	.46	1.78

TABLE 2

NUMBER OF TIMES THE CARTOONS WERE SELECTED AS PREFERRED, AND ASSOCIATED MEAN PREFERENCE RANKS FOR THE FOUR LEVELS OF CARTOON DIFFICULTY

Grade	N	Cartoon difficulty level							
		Easy		Moderately difficult		Difficult		Impossible	
		Preference	Rank	Preference	Rank	Preference	Rank	Preference	Rank
3rd	20	20	2.18	22	1.98	11	2.85	7	2.95
5th	20	19	2.15	24	1.85	9	3.00	8	3.00
7th	20	12	2.65	22	2.00	16	2.48	10	2.88
Total	60	51	2.33	68	1.94	36	2.77	25	2.94

each of three categories, these three categories would each receive a rank of 2, and the remaining category, a rank of 4. The means of these rankings are presented in Table 2. Friedman's analysis of variance of ranks test indicated that the children's preferences differed significantly as a function of difficulty level ($F_{2.03/119.77} = 8.48$; $p < .001$).

Discussion

A phenomenon appears to have been isolated in this study that has not heretofore been demonstrated in humor research. The magnitude of the mirth response was found to depend upon the degree of cognitive congruence existing between the cognitive demand features of the humor stimulus and the cognitive resources of the individual. It is at the point where comprehending the joke taxes the individual's cognitive structure that the humor response is maximal. Consistent with this thesis and certain data of an earlier study (Zigler et al., 1966a), the findings revealed that complete and easy comprehension does not result in the greatest amount of laughter. It is in the intermediate range of difficulty that children obtained both the highest mirth response and preference scores. This phenomenon was found at each of the three grade levels, and in both sexes. Furthermore, the striking parallel between the mirth response and the preference findings indicated that the subject's pleasure could be assessed equally well by recording either his spontaneous laughter or requesting a considered judgment of the cartoons.

The findings of the present study are not only consistent with White's effectance notion, but are in keeping with a variety of theoretical views and everyday observations that have emphasized that children enjoy most that which lies at the growing edge of their capacities. The value of the present study would appear to lie in its clear experimental demonstration of this phenomenon, and in the evidence it provides that humor partakes of a phenomenon that has been thought to characterize other more obviously cognitive behaviors. That humor has many features in common with more recognized cognitive creative acts has recently been emphasized by Koestler (1964). On the other hand, it should be clear that there are determinants other than the cognitive-challenge factor which influence the magnitude of the mirth response. The importance of the content of the humor stimulus interacting with the psychodynamic features of the individual has already been noted. Neither the cognitive challenge, nor the psychodynamic factor appears capable of explaining why incomprehensible cartoons evoke as much mirth as was found in this and in earlier studies (Zigler et al., 1966a, 1966b). What appears to happen in many instances when the child is unable to comprehend the joke correctly is that he sees it as funny for reasons other than the correct one. Often he focuses on a particular feature of the cartoon and finds it uproariously funny. Much of this behavior is probably precipitated by the set involved in the instructions of the task. Once he is told he is to see cartoons and queried about their funniness, the child is confronted with the problem of deducing just why the cartoon is funny. Confronted with such a problem, the child does the very best he can.

Some further speculations concerning the mechanisms mediating the cognitive challenge phenomenon appear to be in order. This study was formulated within a framework that utilized White's concept of effectance. The notion here was that the gratification of the effectance motive occurs when the child uses his cognitive ability to his fullest potential, and that laughter was the emotional expression of this gratification in the same way that satiety would accompany the gratification of the hunger motive. The humor stimuli inducing gratification of the effectance motive should take on reinforcing properties, and thus be preferred by the child when asked to give a preference.

Although this explanation would appear to be consistent with the findings of the present study, a second hypothesis should be entertained. The cognitive-congruence phenomenon found in the present study does not require the type of motivational foundation which inheres in the current thinking on effectance or mastery. One can utilize a motivation-free information-processing orientation and assert that the intrinsic nature of the individual is such that he prefers or enjoys those stimuli which are at a complexity level congruent with his cognitive structure or "channel capacity" (Munzinger & Kessen, 1964). This interpretation is consistent with recent findings obtained on individuals at various age levels (Brennan, Ames, & Moore, 1966; Munzinger & Kessen, 1964).

The distinction between the two hypotheses, though somewhat subtle, is a meaningful one and would appear to generate differential predictions. The effectance hypothesis would generate the prediction that the humor response would depend on two factors: (*a*) the degree of congruence between the complexity of the humor stimulus and the complexity of the observer's cognitive structure, and (*b*) the strength of the observer's motive after mastery. The information-processing hypothesis generates the expectation that the magnitude of the humor response and/or preference responses would depend on the first factor alone. It is, of course, possible to test children who are matched in respect to their cognitive complexity (e.g., MA), but who differ on some independent assessment of the motive to be intellectually successful (e.g., underachievers versus overachievers). Such a study would aid in determining which of the two hypotheses discussed is the most appropriate one for encompassing the clear findings of the present study.

REFERENCES

BRENNAN, W. M., AMES, E. W., & MOORE, R. W. Age difference in infants' attention to patterns of different complexities. *Science,* 1966, **151**, 354–356.

FLUGEL, J. C. Humor and laughter. In G. Lindzey (Ed.), *Handbook of social psychology,* 1954, **2**, 709–734.

FREUD, S. Jokes and their relation to the unconscious. *Standard edition.* Vol. 8. London: Hogarth Press, 1960.

KOESTLER, A. *Act of creation.* New York: Macmillan, 1964.

LEVINE, J. Humor. In, *International encyclopedia of the social sciences.* New York: Macmillan, in press.

MUNZINGER, H., & KESSEN, W. Uncertainty, structure and preference. *Psychological Monographs,* 1964, **78**(9, Whole No. 586).

WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychological Review,* 1959, **66**, 297–333.

ZIGLER, E., LEVINE, J., & GOULD, L. Cognitive processes in the development of children's appreciation of humor. *Child Development,* 1966, **37**, 507–518. (a)

ZIGLER, E., LEVINE, J., & GOULD, L. The humor response of normal, institutionalized retarded, and noninstitutionalized retarded children. *American Journal of Mental Deficiency,* 1966, **71**, 472–480. (b)

(Received June 28, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 337–341

PERFORMANCE UNDER SOCIAL REINFORCEMENT AS A FUNCTION OF MASCULINITY-FEMININITY OF EXPERIMENTER AND SUBJECT[1]

THOMAS R. SHULTZ[2] AND WILLARD W. HARTUP

University of Minnesota

The effects of sex of E, sex typing of E, and sex typing of S on performance under social reinforcement were investigated. Es were selected on the basis of MMPI Mf scores and included 2 masculine males, 2 feminine males, 2 masculine females, and 2 feminine females. 5 masculine male Ss and 5 feminine male Ss (also selected by MMPI Mf scores) were tested by each E in a simple sorting task. The main dependent variable was change in rate of response from a 1-minute base-line period to a 6-minute period during which E delivered supportive statements. Analysis revealed that Ss tested by masculine male Es had higher difference scores than Ss tested by feminine male Es, masculine female Es, or feminine female Es. Rate of response declined over time more for feminine Ss than masculine Ss with the greatest decrement occurring for feminine Ss tested by female Es. The findings are discussed in terms of anxiety reduction, an interpretation consistent with the results of postsession ratings of E by S.

A number of recent studies have shown that performance on simple tasks under social reinforcement is affected by the subject's sex membership and the sex membership of the experimenter (Stevenson, 1965). A "cross-sex" effect has most frequently been reported; that is, male subjects are more effectively reinforced by female experimenters than by male experimenters, while female subjects are more effectively reinforced by male than by female experimenters. These effects have been replicated across subject populations extending from preschool children to college students. Although the literature in which cross-sex effects are reported is extensive, some studies have failed to find these effects at certain ages (e.g., Stevenson, 1961).

The nature of the mechanisms underlying the effects of sex membership on socially reinforced performance is unclear. One possibility is that the effects derive from the social experiences producing sex typing in the individual. For example, there may be a perceived deprivation of opposite-sex support during childhood and adolescence that enhances the incentive value of approval from opposite-sex experimenters. Alternatively, resolution of the Oedipus conflict may produce attraction to persons of the opposite sex, leading the individual to perform particularly well for experimenters of the opposite sex.

The strategy of all the studies revealing the cross-sex effect has involved randomizing individual differences among subjects and experimenters in degree of sex typing (i.e., the extent to which the person is identified with the male or female sex role). If, however, mechanisms such as those cited above underlie the cross-sex effect, it can be hypothesized that this effect should be evident only when the subjects are normally sex typed and the experimenters are representative of their own gender, both in terms of physical appearance and behavior.

The primary purpose of this study was to explore the relation of sex typing (masculinity-femininity) of both subject and experimenter to performance in a socially reinforced task. The study involved male subjects only, selected according to their scores on a standard *Mf* scale. *Both* male and female experimenters were employed, who were also selected on the basis of *Mf* scores. The study thus permitted an analysis of sex-typing factors as these interact with sex of the experimenter in determining performance under social reinforcement.

METHOD

Subjects

The pool from which subjects for this study were drawn consisted of 130 male volunteers from an introductory psychology class of approximately 3,000 students at the University of Minnesota. This study was one of many being conducted simultaneously with students in this class. The 60 items of the MMPI *Mf* scale plus 30 buffer items were administered to the persons in this pool. Test administration was conducted in two groups. Two men scoring at the extreme masculine end of the distribution and two scoring at the feminine extreme were invited

---

[1] The research reported in this paper was supported by a grant from the John E. Anderson Research Fund.

[2] Now at Yale University.

to serve as experimenters for the study (see *Experimenters* below). From the remaining 126 men, the 40 most masculine-scoring and the 40 most feminine-scoring men constituted the two experimental groups of subjects for the study (masculine and feminine). The *Mf* scores of the masculine subjects ranged between raw scores of 20 and 25, with a mean of 22.9 and standard deviation of 1.8. *Mf* scores for the feminine subjects ranged between 28 and 36; the mean for this group was 31.2 and the standard deviation was 2.2.

Experimenters

Four male experimenters were selected as described in the preceding section. Two were classified as feminine experimenters on the basis of *Mf* scores of 36 and 37. The other two were classified as masculine experimenters, both scoring 20 on the *Mf* scale. A check prior to running the experimental sessions revealed that the experimenters did not know any of their subjects personally. The male experimenters were between 19 and 23 years of age. Major differences in dress and grooming were not evident between the masculine and feminine male experimenters.

Four female experimenters were recruited from two introductory child psychology classes. *Mf* scores for these women were available from testing at college entrance. Two were classified as feminine on the basis of raw scores of 45 and 46 (the two most feminine-scoring women in these classes). The other two, classified as masculine, had raw scores of 27 and 30 (again, the extreme scores among the women in these classes). The four female experimenters were between 20 and 22 years of age and, like the men, did not evidence major differences in dress and appearance.

Apparatus

The task used for measuring the effects of social reinforcement involved a marble-dropping apparatus described by Stevenson & Fahel (1961). Briefly, it consisted of a table with two bins. One, located at the subject's left, contained some 1,500 marbles of six different colors. The other bin, located at the subject's right, was covered with a plate containing six randomly placed holes with edges colored correspondingly to the marbles. A partition, approximately 6 inches high and traversing the length of the table, served as a shield between the experimenter and the subject. In front of the experimenter, and shielded from the subject, was a green light. This light was activated by a concealed observer in order to signal the time for the experimenter to dispense reinforcement and the end of the task. An electronic device served to count each marble as the subject dropped it through a hole into the bin.

Procedure

Each subject came, by appointment, to a waiting room where one of the investigators introduced him to his experimenter. The experimenter then escorted the subject to the experimental room, both took their seats at the apparatus, and the experimenter read the following instructions:

> This is a sorting task. These marbles are of different colors and they go in the hole of the same color in this panel. Pick up one marble at a time and drop it in a hole. I'll tell you when to stop. OK. Begin.

The first minute of the task was used to establish the subject's base-line rate of performance. The experimenter was instructed to attend to the signal light and to deliver no verbal reinforcement to the subject during the base-line period. At the beginning of the second minute, the experimenter was signaled to deliver one of six supportive statements. The statements, such as, "That's very good," "That's fine," "You're doing very well," were typed on a card, in one of three randomized orders, and placed on the panel in front of the experimenter. The experimenter was signaled to deliver one of the statements every 30 seconds for 6 minutes. The experimenters were trained not to respond to overtures from the subject, but only to repeat relevant parts of the instructions or to delay questions until later. The experimenter was also instructed to look at the subject when delivering a supportive comment.

Neither experimenters nor subjects were advised of the purpose of the study. Training of the experimenters was accomplished by explaining the apparatus and the procedure and providing each with an opportunity to practice several times while being observed by one of the investigators. The experimenters were told that they would be observed from behind a one-way window during all of their sessions with the subjects.

Each of the eight experimenters tested five masculine and five feminine subjects, assigned randomly. Resulting from this procedure was a 2 (Sex Typing of Subject) × 2 (Sex of Experimenter) × 2 (Sex Typing of Experimenter) design.

Postsession Ratings

Immediately following the marble-dropping task each subject rated his experimenter on a series of 7-point scales. The ratings were completed in another room and the experimenters were not aware that they were being rated.

The ratings were modeled on the semantic differential (Osgood, Suci, & Tannenbaum, 1957) and consisted of 42 pairs of polar adjectives related to masculine and feminine characteristics. One adjective pair included was actually "masculine-feminine." Each of the remaining 41 pairs included a word which Gough (1952) employed in an adjective check list and which differentiated high-feminine from low-feminine subjects. The other member in each of these 41 pairs was chosen by consulting *Roget's Thesaurus*. The adjective pairs were presented in list form with the order of the words in each pair randomized.[3]

---

[3] A three-page sample of the rating scales used in this study has been deposited with the American Documentation Institute. Order Document No. 9324.

TABLE 1

MEAN DIFFERENCE SCORES FOR MASCULINE AND FEMININE MALE SUBJECTS REINFORCED BY MASCULINE AND FEMININE MALE AND FEMALE EXPERIMENTERS

	Min.					
	1	2	3	4	5	6
Masculine male *S*						
Masculine male *E*	2.4	3.3	2.4	3.9	3.1	1.0
Feminine male *E*	−0.8	−0.8	−2.3	−1.6	−2.6	−1.8
Masculine female *E*	0.5	−2.3	−0.8	−1.5	−1.9	−2.4
Feminine female *E*	−1.2	−0.4	−1.6	0.8	−0.7	−1.3
Feminine male *S*						
Masculine male *E*	1.2	0.7	1.6	1.0	0.4	0.6
Feminine male *E*	−1.0	0.7	0.5	−1.9	0.3	−1.6
Masculine female *E*	−0.8	−0.6	−1.5	−4.0	−5.2	−4.1
Feminine female *E*	−1.8	−0.3	−2.3	−1.4	−3.3	−4.0

RESULTS

Base-Rate Scores

An analysis of variance was conducted on the base-rate scores defined as the number of marbles dropped during the first minute of the task when no verbal approval was dispensed. Factors included in this analysis were sex typing of subject, sex of experimenter, and sex typing of experimenter. None of the main or interaction effects were significant in this analysis.

Difference Scores

Six difference scores were computed for each subject by subtracting the number of marbles dropped during the base-rate period from the number dropped during each subsequent minute. Mean difference scores for each of the eight subgroups are presented in Table 1. The data were subjected to an analysis of variance including minutes in the session as a within-subjects factor in addition to those factors listed above. The difference scores of subjects tested by men were significantly greater than the scores of subjects tested by women ($F = 5.91$, $df = 1/72$, $p < .05$). There was an interaction of borderline significance between sex typing of experimenter and sex of experimenter ($F = 3.96$, $df = 1/72$, $p < .06$), indicating that the subjects tested by masculine male experimenters had higher difference scores than subjects tested by feminine male experimenters, masculine female experimenters, or feminine female experimenters. The superiority of subjects tested by masculine male experimenters increased over time in the session as revealed by a significant interaction between sex of experimenter, sex typing of experimenter, and minutes ($F = 3.01$, $df = 5/360$, $p < .05$). The preceding results are presented in Figure 1, which shows the mean difference, plotted for each 1-

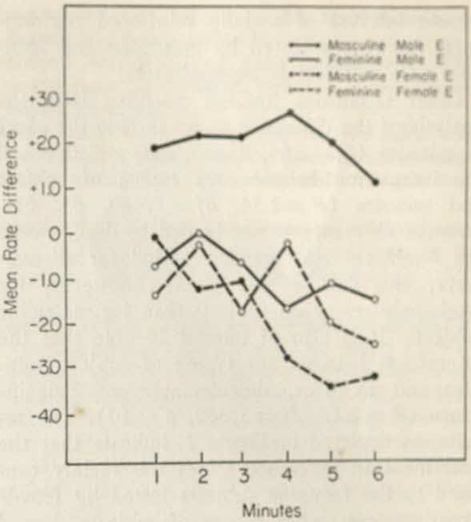


FIG. 1. Mean rate difference scores for six 1-minute intervals for male subjects tested by masculine and feminine male and female experimenters.

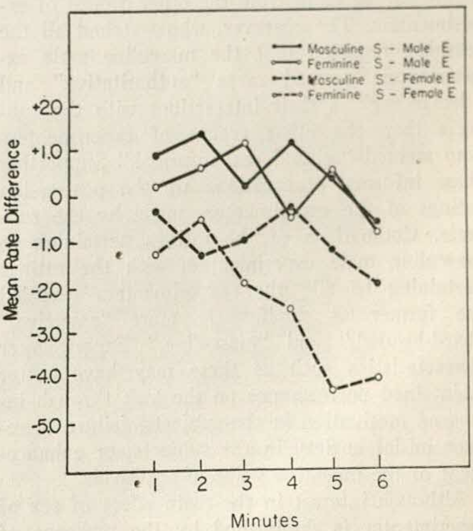


FIG. 2. Mean rate difference scores for six 1-minute intervals for masculine and feminine male subjects tested by male and female experimenters.

---

from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D. C., 20540. Remit in advance $1.25 for microfilm or $1.25 for photocopies and make checks payable to: Chief, Photoduplication Service, Library of Congress.

minute interval of socially reinforced performance, for groups tested by masculine and feminine male and female experimenters.

Other significant findings emerging from the analysis of the difference scores include the effect of minutes ($F = 5.84$, $df = 5/360$, $p < .05$) and the interaction between sex typing of subject and minutes ($F = 2.54$, $df = 5/360$, $p < .05$). Average difference scores tended to decline over the 6-minute reinforcement period for all subjects; the decline was greater, however, for femininely sex-typed subjects than for masculine subjects. It is also of interest to note that the interaction between sex typing of subject, minutes, and sex of experimenter approached significance ($F = 2.12$, $df = 5/360$, $p < .07$). The results, as depicted in Figure 2, indicate that the decrement in difference scores was mainly confined to the feminine subjects tested by female experimenters; other groups of subjects showed little or no decline.

DISCUSSION

The results indicate that sex typing of both the experimenter and the subject is related to performance on socially-reinforced tasks. First, the most efficacious reinforcing agent for the male subjects in this study was a masculine male. Although all the experimenters employed in the study followed a standard procedure with their subjects, the possibility remains that the effectiveness of the masculine male experimenters derived from subtle behavioral characteristics distinguishing them from the other groups of experimenters. The observer, who watched all the sessions, reported that the masculine male experimenters seemed more "authoritative" and "demanding" in their interactions with the subjects than the other groups of experimenters who seemed "relaxed and informal." Supporting these informal observations are the postsession ratings of the experimenters made by the subjects. Comparison of the ratings pertaining to masculine male experimenters with the ratings pertaining to all other experimenters revealed the former as significantly more "assertive," "hard-headed," and "masculine." Experimenter characteristics such as these may have better maintained performance on the task through increased motivation in the subjects, failure to reduce initial anxiety in the subjects, or enhancement of the incentive value of approval.

Although interest in the main effect of sex of experimenter is diminished by the presence of significant interactions involving this factor, it should be noted that the performance of subjects tested by men was superior to the performance of subjects tested by women. This finding is discrepant with the results of those earlier studies reporting a cross-sex effect. It should be remembered that the selection of both experimenters and subjects in the present study was carried out quite differently from the procedures used by other investigators. Participants in the present study were selected from the extremes of a distribution of *Mf* scores, rather than randomly with respect to *Mf*. Failure to replicate the cross-sex effect in this instance suggests that this effect may not be stable across groups at all levels of sex typing. It is interesting to note that the subjects tested by men in this study rated their experimenters, in contrast to subjects tested by women, as significantly more "frank," "violent," "calloused," "harsh," "hard-headed," "hard-hearted," "dejected," and "masculine." The picture of the efficacious male experimenter described by these differences is largely one of hardness and assertiveness, findings which are consistent with the characteristics ascribed to the most effective *subgroup* of experimenters, the masculine males.

With respect to the subject's own masculinity or femininity, the most striking results include a greater decline in performance, over time, by the feminine subjects relative to the masculine subjects. The results of previous studies indicate that feminine males manifest greater anxiety than masculine males (Dahlstrom & Welsh, 1960; Sutton-Smith & Rosenberg, 1960). The results of other studies have shown that supportive comments sometimes serve to reduce anxiety in the high-anxious subject, hence reducing level of motivation and rate of response in tasks involving social reinforcement (Ruebush & Stevenson, 1964; Stevenson & Hill, 1965; Stevenson, Keen, & Knights, 1963). Thus, the poorer performance of the feminine male subjects in the present study as contrasted with the masculine males could have resulted from greater anxiety reduction and reduced task motivation by the experimenter's verbal approval in the feminine subjects.

It is interesting that among the feminine subjects the decrease in mean response rate over time was greater for those subjects tested by women than those tested by men. The social learning histories of feminine men probably include greater contact and more frequent support from females than is the case with masculine men (Terman & Miles, 1936). Consequently, approval from females may have particularly great anxiety-reducing potential for feminine male subjects, thereby producing sharply decreasing scores in this subgroup.

The difference in performance between the masculine and feminine subjects in this study was paralleled by differences in the way these two groups rated their experimenters. Masculine subjects rated their experimenters more positively than did the feminine subjects (as more "ambitious," "selfless," "unwary," "frank," and "clever"). It can be argued that in these ratings the feminine male subjects projected less positive self-concepts than the masculine male subjects. To the extent that anxiety is a correlate of negative self-concept, as has been found by Lipsitt (1958) and Mitchell (1959), the differences in postsession ratings of their experimenters made by feminine and masculine subjects support the explanation (presented above) of the performance results associated with sex typing of the subject.

The generality of the present findings is limited by the small number of experimenters employed. Nevertheless, the study suggests the importance of sex typing (or factors associated with it) as an influence on socially reinforced performance.

REFERENCES

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook.* Minneapolis: University of Minnesota Press, 1960.

GOUGH, H. G. Identifying psychological femininity. *Educational and Psychological Measurement,* 1952, 12, 427–439.

LIPSITT, L. P. A self-concept scale for children and its relationship to the Children's Form of the Manifest Anxiety Scale. *Child Development,* 1958, 29, 463–472.

MITCHELL, J. V. Goal-setting behavior as a function of self-acceptance, over- and under-achievement, and related personality variables. *Journal of Educational Psychology,* 1959, 50, 93–104.

OSGOOD, C. E., SUCI, G., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univeristy of Illinois Press, 1957.

RUEBUSH, B. K., & STEVENSON, H. W. Effects of mothers and strangers on the performance of anxious and defensive children. *Journal of Personality,* 1964, 32, 587–600.

STEVENSON, H. W. Social reinforcement with children as a function of CA, sex of E, and sex of S. *Journal of Abnormal and Social Psychology,* 1961, 63, 147–154.

STEVENSON, H. W. Social reinforcement of children's behavior. In L. P. Lipsitt & C. C. Spiker (Eds.), *Advances in child development and behavior.* Vol. 2. New York: Academic Press, 1965. Pp. 97–126.

STEVENSON, H. W., & FAHEL, L. The effect of social reinforcement on the performance of institutionalized and noninstitutionalized normal and feebleminded children. *Journal of Personality,* 1961, 29, 136–147.

STEVENSON, H. W., & HILL, K. T. The effects of social reinforcement and nonreinforcement following success and failure. *Journal of Personality,* 1965, 33, 418–427.

STEVENSON, H. W., KEEN, R., & KNIGHTS, R. M. Parents and strangers as reinforcing agents for children's performance. *Journal of Abnormal and Social Psychology,* 1963, 67, 183–186.

SUTTON-SMITH, B., & ROSENBERG, B. G. Manifest anxiety and game preferences in children. *Child Development,* 1960, 31, 307–311.

TERMAN, L. M., & MILES, C. C. *Sex and personality: Studies in masculinity and femininity.* New York: McGraw-Hill, 1936.

(Received June 28, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 341–349

THE PHONEMIC CLAUSE AS A UNIT OF SPEECH DECODING

ALLEN T. DITTMANN AND LYNN G. LLEWELLYN

Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland

It is argued here that spoken language is decoded by the listener in word groups called phonemic clauses. Behavioral data, the location of listener responses in relation to the speaker's phonemic clauses, provide the evidence. Pairs of Ss spoke to each other alternately on topics of mutual interest in a situation resembling that of a telephone conversation. The listeners were instructed to let the speakers know that they were listening and interested. Their responses were found to be located almost exclusively at the ends of the speakers' phonemic clauses rather than within them, whether or not there were hesitations at either point. Further, only those types of clause ending which sound most "final" were followed by listener responses.

In this paper we present indirect but strong circumstantial evidence that the phonemic clause is a basic unit which the listener uses in decoding speech. The unit was originally described by Trager and Smith (1951) as necessary to the linguistic analysis of spoken language, applicable to groups of speech sounds larger than the phoneme and morpheme. Boomer and Dittmann

(1962) found that thresholds for recognizing hesitations are markedly affected by their locations in relation to phonemic clauses. In a further study of hesitation forms and phonemic clauses, Boomer (1965) argued that this unit is fundamental to the grammatical encoding process in speech.

Boomer hypothesized that hesitations are indicators of decision processes as the speaker formulates what he is going to say next and how he is going to say it. Boomer traced his hypothesis to a related one by Lounsbury (1954), who reasoned that hesitations would occur at points of high statistical uncertainty and thus mark the beginning of units of encoding. While Lounsbury did not specify the unit at which these points might be found—in fact he said, "units of any given order [p. 99]"—he did note that "The observations which lead us to formulate this hypothesis have been focused on the sequencing of words [p. 99]." He then went on to propose studies on word transition probabilities as the relevant testing ground for his hypothesis. Boomer used the phonemic clause rather than the word as the unit, 3 minutes of uninterrupted speech from each of 16 subjects as the corpus, and found that hesitations were located predominantly at the beginnings of clauses, where syntactic and lexical decisions may not yet have been made. The statistical uncertainty of early words in phonemic clauses, incidentally, is likely to be low.

This unit thus appeared to be established as a promising one for further work in speech encoding. For both logical and empirical reasons the next step was to look for evidence that the same unit is important in decoding as well: the logical reason is that encoding and decoding processes must be closely related if a communication system is to operate efficiently. Empirically, there are compelling results from experiments by Miller and his colleagues that units larger than the individual word are the functional base of speech perception; for an overview see Miller (1962). That work used sentences which were composed for specific experimental conditions instead of conversational speech, and the units were based on considerations derived from the transformational grammar of Chomsky (1957). A possible line of evidence from a more naturalistic setting became apparent as we were analyzing data for another experiment which used the phonemic clause as the unit of analysis. The protocols for that study were 15-minute interviews with college students in which one of us (ATD) served as the interviewer. These interviews were designed to be as unstressful for the subjects as we could make them, and the interviewer interposed many comments and "um-hmm's," by which he hoped to keep the conversation going smoothly and show the subjects that he was attentive and interested in what they had to say. In the course of examining these protocols we noticed that the spontaneous vocal listening responses of the interviewer seemed to be inserted almost exclusively at the boundary points, called junctures, between the speakers' phonemic clauses and almost never at any points within the clauses. Since only one listener's behavior had been examined, a listener who was very much involved in phonemic clause analysis, we then turned to material representing two other interviewers. The same pattern emerged. In this preliminary work we studied a total of 675 phonemic clauses (which ranged from 2 to 10 words, with a mode of 5), in which the listener responded 240 times at junctures and 25 times at other points in the speech.

The phonemic clause, then, is a unit which influences speakers' behaviors as Boomer had found, and listeners' behaviors as well. Boomer related the speech behaviors he studied, and by implication the unit itself, to encoding in speech, as had Lounsbury and others before him, and we have in parallel way related the listener behaviors we have seen in our preliminary results to decoding. Before continuing on to any further data, we shall need to explain our conception of the decoding process, since only then will the use of listener responses as evidence make sense.

While speech is composed of sounds, only combinations of these sounds are ever heard in conversation, or words in the usual sense of that term. Words vary greatly in their information value, as an examination of any language will show: some recur so often and in so many different contexts that it is clear that they cannot convey much information, while others occur so seldom that their appearance marks something quite new. The very frequent words are usually referred to as function words, since they serve to relate the other words, termed lexical items, to one another. In any one conversation the function words will be the same as in any other conversation; they are needed no matter what is being discussed. The lexical items, on the other hand, vary according to the subject matter under discussion.

In an ingenious experiment Berry (1953) showed that these information differences are reflected in the way the words are said. Using recordings of telephone conversations, he marked the transcript to indicate the stress received by each word (strong, secondary, and unstressed), then prepared a word-frequency distribution of the corpus. He found the familiar curve of decreasing frequency with rank for the unstressed words and a complementary reversed curve for the stressed words. Words of secondary stress were distributed in a flat, bell-shaped curve. Uncommon words, the high-information lexical items, received strong stress, while function words were unstressed. Those of medium information were sometimes stressed and sometimes not.

A phonemic clause must have one strongly stressed word, and it may have other words of secondary stress. Degree of stress is relative within each clause, so that until the entire clause has been uttered the listener cannot identify the word which will receive the primary stress, and therefore convey the greatest information. Thus it is not until the end of the clause that he can tell what the main lexical item of the clause has been, relate it to others he has heard, and listen for what may follow. In the view presented here, the words of the clause cannot be decoded until the main lexical item has been identified. Relating one to another as the series of clauses unfolds in an entire utterance is a more complicated matter, not being studied here, probably best described as "understanding" or "comprehension."[1]

By now our conception of the decoding process was clear. The listener must wait until an entire phonemic clause has been uttered before he can discriminate the main lexical item from the other material in the group. Until he has made that discrimination he has not, in effect, heard what the speaker has said during that time. A major clue that a clause has come to an end lies in the rhythmic pattern of stress and pitch in the group of words and the juncture that marks its end. Our hypothesis was that listening responses will occur predominantly following junctures at the ends of phonemic clauses and seldom at any point within a clause.[2] Although analysis of the preliminary material lent strong support to this conception, it raised two specific questions which could only be settled by a more formal study:

1. A simpler explanation of our preliminary results might be that listeners, not wishing to interrupt the speaker, respond at pauses wherever they may occur and not selectively at junctures. According to Boomer's study cited above, about half of all pauses are juncture pauses. These are significantly longer than other pauses and might therefore be considered more conducive to polite response. The difference Boomer found (1.027 seconds for juncture pauses, .747 second for hesitation pauses) is not great enough to account in full for the nearly 10:1 difference in response locations which we had already found. Nevertheless, some more precise way of comparing these two alternative hypotheses was certainly called for.

2. Since we now had a hypothesis firmly in mind, our subsequent work of locating junctures in speech was subject to a strong contaminating influence. In spite of the high reliability of this procedure (see Dittmann & Wynne, 1961) all previous work—including that of the reliability study itself—had been done before this hypothesis had been formulated. Now the person locating junctures in a speaker's material had the additional cues of the listener's responses, which are recorded on the same tape, thus opening the way for spurious confirmation of the hypothesis.

The form of the data for the present study was therefore dictated by these considerations: it must include a record of time so that pause length could be measured, and the sound signal from the speaker and the listener must be recorded independently so that junctures could be located without contamination.

While listening to the material in our preliminary data, another idea emerged. There appeared to be systematic selection of the junctures at which the responses occurred. It should be recalled that there were listener responses to about one-third of the junctures we examined. It was our impression that these were more likely to be the most "final" sounding of the three types of juncture. The protocols studied so far had been coded only for presence or absence of junctures, not for type of juncture also, so more completely coded material was needed to check this impression. There was a possibility, too, of interaction with the alternative hypothesis that listeners respond at pauses, since the type of juncture most often responded to might also have longer pauses associated with it. Thus it was doubly important that the data for the more formal study be coded for type of juncture as well as for their presence or absence.

---

[1] In this line of reasoning the term decoding is restricted to one level of analysis or unit, the word. In spite of unclarity on some points, there is general agreement that decoding must occur with respect to smaller units as well, such as phonemes and morphemes, but that substituting meanings for sounds, or semantic decoding, begins at the level of the word (see Osgood, 1963, p. 742).

[2] Vocalizations are not, of course, the only sort of listener response, though they have been capitalized upon in the work reported here. Other work is currently under way in our laboratory on one of the more visible listener responses, the head nod, and its relationship to the vocal response.

A related question had to do with type of listener response. We reasoned that brief vocal sounds like "um-hmm" might be functionally different from more extended comments and questions and give more information about the nature of the decoding process. More time might be required to formulate an extended response, so that these would more likely be found after longer pauses. In addition, the more final-sounding junctures might elicit comments, the others being followed more often by brief vocal sounds. Thus we wished to separate these listener responses in our analysis so that we would be able to make the necessary comparisons with pause length and juncture type.

METHOD

Subjects

The subjects were six college-student volunteers, four males and two females who were serving as normal controls during a 3-month stay at the Clinical Center of the National Institutes of Health. All subjects were acquainted with one another before the experiment.

Procedure

The experimental conditions were designed to simulate an everyday two-person telephone conversation. Subjects were seen in pairs and told that the experiment had to do with how people talk and how they listen. Each subject was instructed to talk to his partner for approximately 2 minutes, telling him about something that had happened during the course of the day, about a movie he had seen, or about any of a number of similar topics. The person listening was asked to respond as he ordinarily would over the telephone, letting his partner know that he was listening by any means he could, asking questions if a point needed clarifying, or in general commenting on the statements made by the speaker. The only restriction was that one person should carry the major portion of the conversation. The subjects were told that at the end of about 2 minutes, they would be asked to reverse roles, the listener becoming the talker, and vice versa.

After receiving instructions the subjects were taken to separate rooms, given hand microphones and sets of headphones, and allowed to familiarize themselves with the equipment. One investigator stayed in one room with one of the subjects, but out of his sight, and designated a time when the conversation was to begin. In an adjoining room with the second subject the other investigator, also out of that subject's sight, recorded the conversation. The speech of the two subjects was recorded on separate channels of a two-channel tape recorder. The subjects' headphones were connected in parallel to both channels so that each could hear his own as well as the other's voice during the session. An oscillograph with chart paper running at .75 inch per second provided a graphic record of the talking sequence which was being recorded simultaneously on the tape. At the conclusion of the conversation the subjects were told the reasoning behind the experiment and the nature of the hypotheses involved.

To control for possible experimenter bias the investigator (ATD) who coded junctures (a procedure described in detail below) neither heard the complete tape recording nor saw the complete typescript of the conversation. He worked with an edited version of the typescript which included only the speech of the primary speaker and listened only to that channel of the tape. The juncture-coded material was then compared with a revised transcript which had been carefully corrected to ensure the inclusion of all nonfluent as well as fluent speech. The revised transcript also contained the exact location of any listener responses during the primary speech segment.

Responses made by the listener during the course of the conversation were classified as either words or brief vocal sounds. The former can be considered a sequence of several words which serve some purpose for the listener like seeking clarification; that is, "How fast do they [machines] go?" The latter, according to Fries (1952), are a means whereby the listener,

> . . . in some inconspicuous but conventional way, gives the speaker signals of this continued attention. In telephone conversations these signals consist of brief oral sounds interjected at irregular intervals but not interrupting the speaker's span of talk. These brief oral sounds are not predictable, but, in a telephone conversation, if such a sequence of utterances occurs without oral signals of attention on the part of the hearer, the speaker usually interrupts with such questions as *do you hear me* or *are you* (*still*) *there* [p. 49].

Examples of this type of response would include utterances like "yes," "um-hmm," "yeah," "I see," "good," "oh," and so on, in order of their frequency in Fries' telephone-conversation data.

The oscillograph strip chart was used to locate and measure all pauses during the conversation. Pauses were designated as either juncture pauses (any pause which follows a juncture) or as hesitation pauses (any pause other than a juncture pause).

Linguistic Variables

Phonemic clause, as used in this study, is the unit of spoken language set forth by Trager and Smith (1951). Identifying these clauses involves paying attention to what are called *suprasegmental phonemes*. Segmental phonemes are the sounds which correspond roughly to the letters of the alphabet as language is written in conventional orthography. The suprasegmentals are those of pitch, stress, and juncture, which are overlaid upon the segmentals to give connected speech its rhythmic characteristic. For the present purposes we shall discuss only stress and terminal juncture, since they are sufficient to define the phonemic clause; pitch contours are necessary to pronunciation, but vary according to other considerations not germane here. As strings of words are pronounced, certain syllables are stressed because of the words in which they occur, and some are given additional stress because of the syntactic structure of the phrase. Periodically a syllable receives the strongest of the stresses, called the primary stress. Between primary stressed syllables others will be heard to receive various degrees of weaker stress, weaker in comparison to that of the primary stressed word in that clause. Also somewhere between the primary stressed syllables will be heard another phenomenon, the terminal juncture. This marks the end of one phonemic clause and is usually found to follow very closely after the primary stress, sometimes immediately after it, sometimes within one (or less often), two, or more syllables. In the ongoing flow of speech, then, primary stresses and terminal junctures appear alternately and serve to demarcate successive phonemic clauses.

In their original work Trager and Smith (1951, p. 49) defined terminal juncture in terms of slight changes of the pitch phoneme at the very end of the final pitch level of the clause. Accordingly they identified three terminal junctures, one in which the pitch is sustained at the very end of the final pitch of the clause, one in which the pitch rises slightly at the very end, and one in which the pitch falls slightly at the very end. In a later paper Trager (1962) noted that terminal junctures have an additional characteristic in English, that of stretching or lengthening the time required to pronounce the final syllables of the clause. The speaker prepares for the juncture, as it were, by slowing down just slightly as he is coming to it. Trager cites some instrumental work by Martin Joos, referred to in a footnote by Hill (1958, p. 24), which shows the effects of this stretching: over the syllables beginning with the primary stressed syllable and ending with the juncture, the average length of one segmental phoneme is added for the sustained juncture, two phonemes for the rising juncture, and three or more phonemes for the falling juncture. Joos' work was not a formal study based upon a large sample, so the numbers should not be taken literally, but the fact of stretching was demonstrated, and the ordering is a useful guide, when added to the final pitch changes, in coding the three types of terminal junctures.

Phonologically defined in this way by primary stress and terminal juncture, the phonemic clause is a unit of the rhythmical patterning of spoken language. The relationship of the unit to the syntax of the language is undoubtedly an intimate one, but is not well understood at this time. The difficulties of conceptualizing the relationship have to do largely with the distinction between the underlying structure of any language and the performance of individuals as they speak the language. The underlying structure dictates both syntactic and rhythmic patterns. In the ideal form there is probably only one combination of these patterns to convey a given message, but in actual performance speakers do not and cannot always coordinate the two perfectly. Sometimes unit boundaries will coincide and sometimes they will not in conversational speech.

Some evidence bearing on this view comes from data which we have collected on coding junctures from the unpunctuated typescript of a 1,000-word speech sample where the coders did not have the sound of the speaker available to them. There were three coders, one of whom knew about phonemic clauses, but had no direct experience coding them; the other two were completely naive. The coders were told that speech, as they had probably noticed, occurs in rhythmical groups of words, like punctuation in writing, and were asked to mark the typescript wherever they thought such a group ended. They reported that they found it most meaningful to read the typescript silently and imagine how the speaker would have grouped the words. Their coding was compared with that of an experienced coder who worked from the audio tape of the passage. More junctures were coded from the sound version of the material than the typescript version: 143 from sound, 118 by one or more coders from transcript. All three coders agreed on 90 of these, or 63% of the junctures identified from the tape. The three coders agreed on only 6 more junctures which were not noted by the coder who heard the speech. Thus 94% of junctures agreed upon from the typescript were also heard in the speech, while only 63% of those heard were consistently picked up from the typescript alone.

This test is not a complete comparison of the relationship between syntactic and rhythmic structures in spoken language. The coders who worked from the typescript were not instructed to make a syntactic analysis of the material, nor were they professionally trained to do so. On the other hand, the typescript gave many clues about performance characteristics other than syntax. Filled pauses, retraces, anacoloutha, and many articulatory difficulties were included in this carefully prepared transcription and should have enabled the coders to guess accurately what the rhythm patterns were. The results obtained from the test indicate that speech rhythm is more than a matter of the words used and their syntactic combinations.

Evaluating the reliability of juncture coding is beset by a number of problems. An earlier report (Dittmann & Wynne, 1961, p. 203) presented percentages of agreement where coding was done by a group on two occasions 6 months apart. Deciding whether to code a juncture or not yielded 93% agreement, while deciding which juncture to code where one was coded both times was somewhat lower, 87%. The main problem lies in the first of these figures. When a juncture has been coded both times the coders have obviously agreed, but can we say they have agreed when no juncture has been coded? If we assume that there is an opportunity for a juncture after every word, we can set up a contingency table and test obtained frequencies against those predicted by the marginal totals. We did this in the earlier study, and the resulting χ^2 was 304.7 ($\varphi = .83$). But this figure is spuriously high, first because we allowed "word" to be defined by the typist as she pressed the space bar, and second because not all words could be followed by a juncture in normal speech, an article which precedes a noun being the most obvious example. If we consider the number of junctures agreed upon on both occasions compared with the total number coded either time, the agreement is 79%, but no statistical test is possible with these figures. In the absence of more precise knowledge about the probability of junctures following different words or groups of words, our estimate of reliability will have to rest here. The other reliability figure, of which juncture to code when there is agreement to code a juncture, can be evaluated more easily: $\chi^2 = 89.5$ ($\varphi' = .63$).

Our method of coding junctures in the present study, we feel, is an improvement over the old method; now we play the tape of the material at three-quarters of its normal speed. The recordings are made at 7.5 inches per second; then for playback a sleeve is placed over the capstan of the tape recorder to increase its circumference by a factor of 1.5, and the machine is run at its "3.75" inches per second speed. In this way the material can be heard much more clearly, especially the stretching phenomenon so important to accurate coding of terminal junctures.

RESULTS

The main hypothesis, that listeners will respond after junctures which terminate phonemic clauses, but rarely respond within clauses, was clearly upheld by the data. A total of 379 phonemic clauses made up the output of the six speakers, and there were 124 listener responses. Of these responses, 110 followed junctures, and only 14 occurred within clauses, a ratio closely resembling that obtained in the preliminary data which led up to the study. The pattern of responding after junctures was followed consistently by all subjects, with one partial exception: 8 of the 14 nonjuncture responses were contributed by one of the subjects, making his ratio of juncture to nonjuncture responses a little over 3:1.

While the main hypothesis seemed firm so far, the results did not completely rule out the alternative hypothesis that listeners respond to pauses as such and not to junctures alone. To test this idea further, frequency of responses at juncture pauses was compared with that of responses at hesitation pauses. The results are presented in Table 1, which shows that responses are located almost exclusively at juncture pauses. Separate analyses for the six individual subjects yielded similar results: chi-squares were significant at the .05 level or beyond for all subjects and consistently in the direction of more responses at juncture pauses.

TABLE 1

RESPONSES TO JUNCTURE PAUSES AND HESITATION PAUSES

	Listening response	No listening response	Total
Juncture pauses	97	98	195
Hesitation pauses	4	119	123
Total	101	217	318

Boomer's finding that juncture pauses were longer than hesitation pauses was considered next as a possible contributing factor to the overall significance of the results. Based solely on the ratio of juncture pause to hesitation pause lengths which Boomer found, approximately 10 responses to junctures would be expected for every 7½ responses to hesitations. In our data, however, no significant difference was found between juncture and hesitation pause lengths, even taking into account the correlation between the two across subjects. It must be noted that our results cannot be compared directly with Boomer's. The differences can probably be attributed to differences in experimental procedure, which will be discussed below, but pause length, in any case, does not seem to be a major contributor to our results.

Earlier we mentioned that a selective factor might be operating with regard to listener response to junctures of different types: those which sounded the most final seemed more often to elicit some form of response. To examine this idea, we compared frequencies of response to the different juncture types. We found that rising and falling types of juncture were responded to significantly more often than sustained junctures ($\chi^2 = 68.06$, $df = 2$; $\varphi' = .424$). No significant difference was found between the response frequencies to the rising and falling types of juncture. Thus the sustained juncture which is the least final sounding of the three types of terminal juncture is contrasted with the other two in rate of listener response.

Next we compared the three types of juncture for the length of pause likely to follow. For this analysis only those junctures which were not responded to could be included, since the listener's response interrupts the pause, and there is no way of knowing how long it might have been if

TABLE 2

RESPONSES TO DIFFERENT PAUSE LENGTHS OF THREE JUNCTURE TYPES

Pause length	Sustained juncture		Rising juncture		Falling juncture		Total		Grand total
	Response	No response	Response	No response	Response	No response	Response	No response	
None	2	119	1	18	10	34	13	171	184
Short	3	13	10	7	39	21	52	41	93
Long	2	6	9	8	34	43	45	57	102
Total	7	138	20	33	83	98	110	269	379

the response had not intervened. Because of the combination of markedly different sample sizes and skewed distributions, the distribution of pause length was trichotomized to yield similar marginal totals (no pause, short and long pauses) and frequency statistics were used. The three juncture types were significantly different on pause length ($\chi^2 = 73.05$, $df = 4$; $\varphi' = .368$). Table 2 shows the distributions: only 14% of the sustained junctures were followed by a measurable pause, as compared with 45% of the rising and 65% of the falling types; where pauses did occur, there was a positive relationship between pause length and juncture type, the sustained juncture being associated with short pauses, and the falling juncture with long pauses.

The results of the data thus far indicate that a possible interaction might have occurred between juncture type, pause length, and listener response, since the most final-sounding juncture was responded to more often, was followed by a pause more often, and was associated with longer pauses. But for the reason noted above, namely, that measurement of pause length is interfered with by responses, it is impossible to include all three variables in the same analysis. Since we recognized that this problem made interpretation of the results difficult, however, we did perform the analysis for whatever trends it might show. The data are presented in Table 2. Information transmission analysis (McGill, 1954) showed that both pause length and juncture type were significantly related to listener response, accounting for 18% and 21% of the total uncertainty, respectively, and corroborating our earlier findings. The interaction term was negative, a possibility in this type of analysis. Using this interaction to hold juncture type constant, the relationship between pause length and listener response is not so pronounced, accounting for only 10% of the total uncertainty, but it is still statistically significant. The combination of difficulty in measuring pause length where there is a response and the low frequencies in some of the cells of our table makes it impossible to be more precise in our conclusions.

Our attention next focused on mode of response: would the listener in a conversation be likely to react with brief vocal sounds or more extended comments and questions, depending upon pause length? Our data do not indicate any such trend. A *t* test for correlated measures using mean pause length before each mode of response did not reach significance. The other relationship to be examined was that between mode of listener response and type of terminal juncture used. A chi-square test was not significant, indicating that simple responses and longer strings of words were as likely to follow sustained junctures as either of the other types. A factor which might temper the obtained negative results on mode of response was a scoring convention: subjects often uttered a response then, after a pause, followed it with a further comment or question. These instances were counted as brief vocal sounds, since that was the immediate response. A much larger body of data would be required to take this scoring problem into account.

DISCUSSION

The results of this investigation strongly support the hypothesis that listeners make use of groups of words as units in decoding speech, groups of the size of the phonemic clause. In addition, the results tell us something more than this: people not only respond to the rhythm of alternating primary stresses and terminal junctures in the speech they are listening to, they also discriminate among different types of junctures, and we know this because they respond differentially to them. Even knowing this, however, we have still not accounted for all the data we have collected. Somewhat fewer than half of the more final-sounding junctures were followed by responses. Why not all of them, or a large majority of them? The subjects were instructed, after all, to respond often. The answer is, of course, that listeners have much more information about when to intervene than simply the rhythm of the speech they are listening to. The content provides many cues, but there are too many possibilities here for study at this time: methods of content analysis would yield units much larger than those needed to answer this question. Syntactic sequences provide many other cues. As we noted earlier, syntax and rhythm patterns are basic structural members underlying the rest of a language and may follow each other exactly in ideally fluent speech. It is our impression, however, based on the close listening required for coding junctures, that there is a good deal of slippage between syntax and performance during conversational speech. Talking is a commonplace activity, but tremendously complicated, and the decision points are many: "Should I put it this way or that way?" "Should I continue with this line of thought at all?" "After this long, drawn-out subject phrase, should the verb be singular or plural?" "Maybe I had better qualify what I just said." "How did I get mixed up with this conditional mode?" "Do I really believe what I am about to say?" These and all the hundreds of other internal questions one has while talking make fluent, "grammatical" speech very difficult. They also make syntactic analysis of ongoing speech just as difficult and lead listeners to wait until they "have really understood" before they respond.

Earlier we mentioned that 8 of the 14 nonjuncture responses in our results were contributed by one individual. It is interesting to note that at least 4 of these responses could be identified as a form of verbal oneupsmanship commonly played by adolescents. This game can best be illustrated by the following sequence: the speaker would say something like, "Do you know . . .," and then immediately be interrupted by this listener with, "No, I don't." The effect of this playfulness, of course, is to dilute the obtained association between terminal junctures and listener responses.

One difference was found between our results and those reported by Boomer (1965), a difference which calls for explanation. He reports (p. 155) that juncture pauses were significantly longer in his subjects' speech than hesitation pauses and, further, that this direction of difference obtained in 15 of his 16 subjects. We, on the other hand, found no significant difference in length between hesitation and juncture pauses, and the direction was evenly divided among our six subjects. In view of the consistency of Boomer's finding, this difference in results does not seem to be a matter of chance variation between the two samples. We believe that the explanation may be found in the comparison of the situations which the two experiments used to obtain the speech samples. In Boomer's study each subject was asked to speak extemporaneously on any topic for 3 minutes, and he was told that the experimenter would not intervene during that time (see Boomer & Dittmann, 1964, p. 325). The experimenter was seated in view of the subject, and, while he did not speak during the 3 minutes, the subject could see that he was attending, and he may well have given some visible cues of his interest. In the present experiment the subjects were in separate rooms, and the listeners were instructed to let the speakers know that they were listening, as in a telephone conversation. As a result there were many interventions, and these, as we have seen, came mainly at junctures. Juncture pauses were thus often terminated before they might have been in a soliloquy, and those which were could not be included in our comparisons of the lengths of the two types of pauses. Thus juncture pause lengths appear shorter in our data than in Boomer's, being about the same length as the hesitation pauses in his data.

REFERENCES

BERRY, J. Some statistical aspects of conversational speech. In W. Jackson (Ed.), *Communication theory*. New York: Academic Press, 1953. Pp. 392–401.

BOOMER, D. S. Hesitation and grammatical encoding. *Language and Speech,* 1965, **8**, 148–158.

BOOMER, D. S., & DITTMANN, A. T. Hesitation pauses and juncture pauses in speech. *Language and Speech,* 1962, **5**, 215–220.

BOOMER, D. S., & DITTMANN, A. T. Speech rate, filled pause, and body movement in interviews. *Journal of Nervous and Mental Disease,* 1964, **139**, 324–327.

CHOMSKY, N. *Syntactic structures*. The Hague: Mouton, 1957.

DITTMANN, A. T., & WYNNE, L. C. Linguistic techniques and the analysis of emotionality in interviews. *Journal of Abnormal and Social Psychology,* 1961, **63**, 201–204.

FRIES, C. C. *The structure of English. An introduction to the construction of English sentences.* New York: Harcourt, Brace, 1952.

HILL, A. A. *Introduction to linguistic structures.* New York: Harcourt, Brace, 1958.

LOUNSBURY, F. G. Transitional probability, linguistic structure, and systems of habit-family hierarchies. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: A survey of theory and research problems*. Baltimore: Waverly Press, 1954. Pp. 93–101.

MCGILL, W. J. Multivariate information transmission. *Psychometrika,* 1954, **19**, 97–116.

MILLER, G. A. Some psychological studies of grammar. *American Psychologist,* 1962, **17**, 748–762.
OSGOOD, C. E. On understanding and creating sentences. *American Psychologist,* 1963, **18**, 735–751.
TRAGER, G. L. Some thoughts on 'juncture.' *Studies in Linguistics,* 1962, **16**, 11–22.
TRAGER, G. L., & SMITH, H. L., JR. *An outline of English structure.* (*Studies in linguistics: Occasional papers, 3*). Norman, Okla.: Battenberg Press, 1951. (Republished: New York: American Council of Learned Societies, 1965)

(Received June 30, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 349–353

BETRAYAL AND RECONCILIATION IN A TWO-PERSON GAME[1]

S. S. KOMORITA AND JOHN MECHLING[2]

Wayne State University

In order to study the development of trust and cooperation, undergraduate psychology students were asked to make choices in a Prisoner's Dilemma game. Under the assumption that the study of distrust would lead to a better understanding of the development of trust, an attempt was made to determine the conditions under which an act of betrayal leads to suspicion, and to determine the variables which influence the process of reconciliation. 3 variables were manipulated: (a) temptation to defect, (b) the amount of harm experienced when betrayed, and (c) the amount of prior cooperation before the betrayal. Although all 3 of the variables had significant effects, the triple interaction and 1 of the 1st-order interactions were also significant.

Of a variety of experimental games suggested by game theory, probably of greatest interest to psychologists is the two-person Prisoner's Dilemma game.[3] This game is structured in such a way that each player is faced with a conflict of motives—a motive to maximize *personal* gain versus the motive to maximize *mutual* gain. Some investigators, therefore, have used it as a paradigm for the study of cooperation and competition. Another interesting aspect of this game is that a cooperative choice implicitly implies trust in the other person, since it would be foolish to cooperate if one did not expect the other to reciprocate such a choice. One of the interesting problems, therefore, is to determine the relevant variables which lead to the development of trust.

The typical paradigm in past research has been to present a large number of trials to subjects and relate certain experimental variables with the degree of cooperative choice over trials. Deutsch, for example, has shown that communication and the motivational orientation of subjects is a critical variable, and Rapoport has shown that the payoff values of the matrix affect the degree of cooperation (Deutsch, 1960; Rapoport & Chammah, 1965).

Rapoport and Chammah (1965) have also shown that the proportion of cooperative choices over a series of trials is U shaped, declining initially, and then increasing on later trials. There are large individual differences, however, and some subjects reach on asymptotic stage where they never make cooperative responses. In such studies where subjects are given a large number of trials it is assumed, explicitly or implicitly, that trust and cooperation is learned and develops over the course of trials.

Although the present study made the same assumption, unlike in previous research an attempt was made to determine the conditions under which an act of betrayal leads to distrust or suspicion, and the variables which influence the process of reconciliation. In effect, the subjects were instructed to assume a cooperative orientation; the cooperative orientation was extinguished by leading them to believe that they had been betrayed; then the recovery of cooperative choice was used as the response measure. It was assumed, of course, that the study of betrayal and reconciliation would lead to a better understanding of the development of trust.

---

[1] This study was supported, in part, by a grant from the Carnegie Foundation to the Center for the Study of Cognitive Processes, Wayne State University.

[2] The junior author was a participant in the National Science Foundation Undergraduate Research Participation Program during the conduct of this study.

[3] For a detailed discussion of this game see Luce and Raiffa (1957), and for a review of empirical research see Rapoport and Orwant (1962) and Gallo and McClintock (1965).

METHOD

Subjects. The subjects for the study were 64 undergraduate students (32 women and 32 men) enrolled in introductory psychology classes, and participating in the experiment to fulfill part of their requirements for the course.

Procedure. All subjects were scheduled in pairs of the same sex and were seated across a table facing each other with a wooden partition between them so that they could not see each other. In front of each subject was a subject control panel displaying the game matrix with two buttons (push-button switches), red and black, corresponding to the two choices in the game. The experimenter sat at right angles to the subjects with another wooden partition separating him from both subjects. A master control panel was wired in such a way that the experimenter could receive information about the subjects' choices on each trial and also permitted the experimenter to send preprogrammed feedback to both subjects simultaneously. When both subjects pressed a button on their panels, a light in one of the cells of the matrix came on indicating to each subject how much each had won as well as the choice made by the other. Another light on the subject control panel served as a signal for the beginning of a trial.

Subjects were instructed that the values of the matrix represented points and that their task was to accumulate as many points as possible. They were also informed that these points would be converted to money at the end of the experiment, not necessarily on a one-to-one basis, but the more points they accumulated the more money they would receive.

In order to insure cooperation at the outset, a modification of Deutsch's cooperative orientation was used (Deutsch, 1960); for example, part of the instructions was as follows:

> Let me emphasize that you should consider yourselves to be partners. You're interested in your partner's welfare as well as your own. You care how she does and she cares how you do. In other words, choose the combination where you can both benefit the most. . . .

After making certain that subjects understood the instruction and the experimental situation, they were instructed not to talk or communicate with each other.

All subjects were led to believe that they were playing with each other, but in actuality they were playing against the experimenter. They were led to believe that they were cooperating with each other for a number of trials, and then were double-crossed on two consecutive trials. Thereafter, the experimenter played "red" (cooperative choice) on three consecutive trials, and subsequently imitated the previous responses of the subject; for example, if the subject played "red," the experimenter played "red" on the next trial. The response measure of reconciliation was the number of trials to reach a criterion of five consecutive cooperative responses after the two betrayals.

TABLE 1

GAME MATRIX WITH TWO INDEPENDENT VARIABLES: TEMPTATION (T) AND LOSS (L)[a]

		Other (*E*)	
		Red	Black
S	Red	5, 5	L, T
	Black	T, L	1, 1

[a] Left and right cell entries denote payoffs to subject and other, respectively.

Independent variables. Three variables, each with two levels, were manipulated. In Table 1 and Figure 1, two of these variables are summarized, where T refers to "temptation" (to defect) and L refers to the loss incurred when betrayed. All subjects received 5 points for the "red, red" combination and 1 point for the "black, black" combination, and as indicated in Figure 1 the manipulated values of temptation were 6 and 9, and the losses when betrayed were 0 and −5.

It was hypothesized that the greater the harm experienced when betrayed, the greater the effects of the betrayal; hence, it was predicted that the −5 condition would require a greater number of trials for reconciliation to occur. With regard to the effects of temptation, it was hypothesized that it would require more trials for reconciliation when temptation was slight than when it was large. The rationale is that when one is betrayed when the other's temptation is negligible, it implies an intent on the part of the other to injure and harm rather than to maximize his own gain.

The third variable consisted of the number of cooperative trials before the betrayal with 4 and 10 trials before the two double crosses. This treatment

T(6), L(0)	
5, 5	0, 6
6, 0	1, 1

T(9), L(0)	
5, 5	0, 9
9, 0	1, 1

T(6), L(−5)	
5, 5	−5, 6
6, −5	1, 1

T(9), L(−5)	
5, 5	−5, 9
9, −5	1, 1

FIG. 1. Game matrices showing values of T (6 and 9) and L (0 and −5).

TABLE 2

PROPORTIONS OF COOPERATIVE CHOICES IMMEDIATELY AFTER THE FIRST OF TWO CONSECUTIVE BETRAYALS

Trials	4 cooperative trials				10 cooperative trials				Mean proportion
	T(6)		T(9)		T(6)		T(9)		
	L(0)	L(−5)	L(0)	L(−5)	L(0)	L(−5)	L(0)	L(−5)	
1	.75	.62	.25	.50	1.0	.87	.75	.87	.70
2	.37	.25	.37	.12	.62	.25	.25	.62	.36
3	.75	.50	.87	.50	.87	.62	.37	.37	.61
4	1.0	.75	.87	.50	1.0	.62	.50	1.0	.78
5	1.0	.50	1.0	.37	1.0	.75	.62	.87	.77

was intended to vary the expectation of a cooperative response. It was hypothesized that if an individual strongly expected a cooperative response from the other person the effects of a disconfirmation of this expectation ought to be greater than if the expectation was not as strong. Accordingly, it was predicted that subjects with 10 mutually cooperative trials would require a larger number of trials for reconciliation than subjects with 4 such trials.

In summary, there were three independent variables, each with two levels: (*a*) temptation, T (6 and 9), (*b*) loss incurred, L (0 and −5), and (*c*) number of cooperative trials before the betrayal (4 and 10 trials). The levels of the three variables were factorially combined to form a 2^3 factorial analysis of variance design with eight subjects in each condition. In addition, there was an equal number of men and women in each condition (2 pairs each), but the analysis indicated that neither main effect nor any of the interactions involving sex was significant; hence, the results for sex differences have been omitted. Finally, a postexperimental questionnaire consisting of a set of evaluative semantic differential scales was also administered to subjects asking them to rate the other person's behavior as well as their own.

RESULTS

The data for five subjects (three in the 10-trial condition and two in the 4-trial condition) were discarded because they responded noncooperatively on one or more trials before the experimenter could double-cross them. For the 64 subjects who responded cooperatively until the first betrayal, Table 2 shows the proportions of cooperative choices for the first five trials after the first of two betrayals. It can be seen that the majority of subjects ignored the first betrayal, but after the second, most subjects reciprocated the defection. In general, the rates of recovery after the two betrayals are very rapid, but this is not unreasonable since the experimenter responded cooperatively for three consecutive trials after the two betrayals. The rates of recovery also seem to vary for the different experimental conditions, and these differences are consistent with the means for number of trials to criterion shown in Table 3. Since the variances of the data in Table 3 were highly heterogeneous, for the purpose of the analysis of variance a reciprocal transformation was used, and Hartley's $F_{\max}$ test indicated that the variances of the transformed data were not significantly different ($F_{\max} = 6.818$; 7, 8; $p > .05$).

Table 4 shows the summary data for the analysis of variance based upon the transformed data. It can be seen that all three main effects are significant at the .05 level. The triple interaction and the Trials × Loss first-order interaction, however, are also significant. It can be seen in Table 3 that these interactions are due mainly to the highly discrepant mean of 22.16 for the high-temptation, high-loss, four-trial condition. Accordingly, in order to provide evidence for the reliability of these interaction effects, this condition was replicated on eight additional subjects.

TABLE 3

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) OF NUMBER OF TRIALS TO CRITERION

	4 cooperative trials		10 cooperative trials		Grand means
	T(6)	T(9)	T(6)	T(9)	
L(0)	6.87 (0.80)	8.75 (5.78)	6.50 (0.57)	9.37 (3.78)	7.87
L(−5)	11.87 (5.41)	22.12 (8.97)	10.37 (5.29)	9.12 (5.69)	13.37
Grand	4 trials: 12.41		10 trials: 8.84		
means	T(6): 8.91		T(9): 12.34		

TABLE 4

SUMMARY ANALYSIS OF VARIANCE TABLE[a]

Source	*MS*	*F*
Temptation (T)	65.205	5.258*
Loss (L)	264.062	21.294**
Trials	72.675	5.861*
T × L	2.891	—
T × Trials	13.142	1.060
L × Trials	112.361	9.061**
T × L × Trials	80.998	6.532*
Error	12.401	

[a] Data for the effects of sex have been omitted since neither main effect nor interactions involving sex were significant.
* $p < .05$, $df = 1/48$.
** $p < .01$, $df = 1/48$.

The mean number of trials to criterion for this group was found to be 24.75, and is consistent with the mean of 22.16 for the original group of eight subjects.

Table 3 shows that when the loss incurred is small ($L = 0$), the effect of the number of prior cooperative trials is negligible. When the loss is large, however, there is a marked effect of the number of cooperative trials when temptation is high but not when it is low. Thus, there seems to be a multiplicative effect of the three variables on trials to criterion.

One possible explanation for the interaction effects is that in the discrepant condition involving high temptation, high loss, and four prior cooperative trials, information about the intentions of the other was more ambiguous. Since perceived intentions were uncertain and the negative consequence of being betrayed was severe, subjects in this condition may have attempted to minimize the risk and consequences of another betrayal. Another explanation is that when temptation and loss are high, a certain number of mutually reinforcing trials are necessary before such reinforcements have any effects. Subjects in the 10-trial condition had experienced the rewards of mutual cooperation for 10 trials, and it is plausible that the 4-trial group had not had a sufficient number of reinforcing experiences to overcome the suspicion and fear of another betrayal.

With regard to the effects of the loss when betrayed, the data indicate, as predicted, that the greater the magnitude of the loss, the more trials it requires for reconciliation to occur. However, the significant interaction also indicates that this effect is greater under the 4-trial than under the 10-trial condition. For temptation to defect, the results indicate that the *smaller* the temptation, the *fewer* trials necessary for reconciliation. This is contrary to the author's prediction and suggests an interesting question concerning the effects of temptation, namely whether the effect of temptation is a function of the subject's *own* temptation to defect, or a function of the subject's fear that the *other* person is tempted. This is especially relevant when the temptation is high, but a similar question may be asked when the temptation is small; that is, when temptation is small, subjects have very little reason to continue responding competitively, but it is plausible that they project this motive to the other person as well.

Still another motivating factor is the amount of resentment or hostility aroused by the betrayal. In the postexperimental questionnaire, subjects were asked to evaluate the other person's behavior as well as their own on a semantic differential scale with bipolar adjectives like "fair-unfair," "honest-dishonest," "clean-dirty," and the data were analyzed for the different conditions. The one significant result was the main effect for temptation ($F = 4.099$; $df = 1, 48$), indicating that the high-temptation groups negatively evaluated the other person's behavior significantly more than the low-temptation groups. Thus, resentment may have played a significant role in the process of reconciliation.

The data for the effects of temptation and consequences of a betrayal are consistent with the results found by Rapoport and Chammah (1965), who used a conventional paradigm where subjects played the game over a large number of trials. Their results showed that the proportion of cooperative choices was smaller when temptation was large and when the negative consequence of making a cooperative response, when the other person defects, was small. Whereas subjects in their study actually played each other over a large number of trials, in the present study there were much fewer trials and all subjects were given a preprogrammed sequence of feedback responses. The results of this study, nevertheless, lend support to the results found by Rapoport and Chammah, and the implications of the results for these two variables seem to be clear—the process of reconciliation seems to follow the same underlying principles as the development of trust.

The implications regarding the effect of the number of prior cooperative trials, on the other hand, are not as clear. Under the assumption that the stronger the expectation, the greater the effects of a disconfirmation of the expectation, it had been predicted that the high-expectation groups would require a larger number of trials for reconciliation. Since the opposite tendency was found for the four-trial group under the high-loss, high-temptation condition, a post hoc explanation is that this group had not experienced or learned the reinforcing effects of mutual cooperation. It is plausible that subjects in this situation, like animals in an avoidance conditioning paradigm, were not "aware" that betrayal (punishment) was no longer forthcoming, and the severity of the loss for the betrayal made it difficult to overcome their fear of another betrayal. The interesting paradox is that if we could somehow coerce or entice subjects to cooperate at the outset, they would experience the mutual benefits of cooperation, and generalizing from the results of this study, trust and cooperative behavior should be greatly facilitated.

REFERENCES

DEUTSCH, M. The effect of motivational orientation upon trust and suspicion. *Human Relations*, 13, 1960, 123–139.

GALLO, P. S., & MCCLINTOCK, C. G. Cooperative and competitive behavior in mixed-motive games. *Journal of Conflict Resolution*, 9, 1965, 68–78.

LUCE, R. D., & RAIFFA, H. *Games and decisions.* New York: Wiley, 1957.

RAPOPORT, A., & ORWANT, C. Experimental games: A review. *Behavioral Science*, 7, 1962, 1–37.

RAPOPORT, A., & CHAMMAH, A. *Prisoner's dilemma: A study in conflict and cooperation.* Ann Arbor: University of Michigan Press, 1965.

(Received June 30, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 353–359

ANXIETY AND REPRESSION AS PREDICTORS OF THE USE OF INCIDENTAL CUES IN PROBLEM SOLVING

GERALD A. MENDELSOHN AND BARBARA B. GRISWOLD[1]

University of California, Berkeley

Prior to solving anagrams, Ss memorized 25 words under interference conditions, another list of 25 words played on a tape recorder. Unknown to Ss, 10 anagram solutions had been in the memorized list (focal incidental cues) and 10 in the interference list (peripheral incidental cues). As predicted, scores on the MMPI R scale (Repression Denial) were negatively related to the use of the cues ($p < .001$). The results were more pronounced for males. Contrary to prediction, A scale (Anxiety) scores were related, for males only, in a curvilinear fashion to cue utilization, the middle A group using the cues least ($p < .01$). All findings were replicated. An interpretation of repression in terms of control of attention and an amendment to Easterbrook's generalization that emotion tends to reduce the range of cue utilization were offered.

It has long been noted by clinicians that the operation of psychological defenses can have important effects on perceptual behavior (see, e.g., Schafer, 1954). While such effects are most evident in the presence of conflict relevant material, recent studies suggest that the use of particular defenses is associated with more general perceptual styles. Holzman and Gardner (1959) reported that female subjects characterized as extreme repressers use the cognitive control of "leveling," and (Gardner, Holzman, Klein, Linton, & Spence, 1959) that isolators tend to be high "scanners." Luborsky, Blinder, and Schimek (1965) similarly found relationships between repression, isolation, and looking behavior. "Isolation," they concluded, "was strikingly associated with venturing to look around more, and the defense of repression with looking around less [p. 270]" as measured by an eye-movement camera. These results are most pronounced for drive-related material but hold, too, for neutral stimuli. These studies gain importance from the fact that they are not beset by the problems of verbal report which have made the perceptual defense literature so difficult to interpret.

The research to be reported is concerned with the relationship of the defense, repression, to the use of incidental and peripheral cues in problem solving. The assumption underlying the experiment is that subjects classified as high repressers

---

[1] The writers wish to express their thanks to Donald W. MacKinnon and the staff of the Institute of Personality Assessment and Research, where this paper was written, for their support and interested criticism. Thanks are due, too, to the Institute of Social Sciences whose grant made the data collection and analysis possible.

will focus attention more narrowly[2] and thus utilize incidental cues less. An additional variable, anxiety, was brought into the study as well, because of its relevance to cue utilization. Easterbrook (1959) concluded, in his review of the effects of emotion on cue utilization, that ". . . emotional arousal acts consistently to reduce the range of cues that the organism uses . . . [p. 183]" and Korchin (1964), addressing himself specifically to anxiety, reached the similar conclusion that the capacity for "dual activity," of which one form is "incidental cognition," is reduced in anxiety. Given the consistency of findings reflected in these conclusions and the presumed relationship of anxiety to defense, it seemed important to the authors to include a measure of anxiety as well as of repression in the study.

Two hypotheses were formulated: (*a*) Strength of repression will be negatively related to the use of incidental cues in problem solving; (*b*) level of anxiety will be negatively related to the use of incidental cues in problem solving. Although interactions between repression and anxiety might be expected, no predictions were made in this regard.

METHOD

Subjects and Independent Variables

The subjects in this study were undergraduates enrolled in the introductory psychology courses at the University of California, Berkeley, for whom participation in an experiment was a course requirement. Data were collected at different times during the academic year on two samples, the first consisting of 33 males and 46 females, and the second, 42 males and 60 females. A 181-item self-report inventory consisting of the first and second factor scales of the MMPI, *A* and *R* (Welsh, 1956), and the Byrne Repression-Sensitization scale (Byrne, 1964) were administered to all subjects. (Since the Byrne scale correlated above .90 with *A*, it was not included in the analyses to be reported.)

The 39 items of the *A* scale tap such factors as doubt, indecision, self-disparagement, dysphoria, sensitivity, negative emotional tone, pessimism, anergia, and disturbance or impairment of thought processes. The 40 items of the *R* scale refer to the denial of aggressive, sexual, and exhibitionistic impulses, lack of social dominance and enjoyment in social situations, concern about health, and avoidance of excitement and emotionality. Welsh (1956) summarized their relationships to other variables as follows:

> It is clear that high A scores are related to disability of a dysthymic and dysphoric nature in which anxiety is prominent. The disorders exhibited by high R scorers are characterized by repression and denial [p. 280].

The most thorough analysis of these two factors is that of Block (1965), who concluded that (*a*) the interpretations of the two factors in terms of acquiescence and social desirability response sets are inaccurate, and (*b*) the labeling of the factors as anxiety and repression may be overextended. "Anxiety," he argued, "is a *state* of the individual where this factor reflects a *characteristic disposition* [p. 110]." Thus, "Welsh's (1956) earlier identification of this factor as 'anxiety' would appear to be improved if changed to 'susceptibility to anxiety' [p. 110]." Block preferred the label "ego-resiliency." Similarly, he noted that the concept repression implies more than "simple impulse suppression" and suggested the label, "ego-control," as more appropriate. However, both the behavioral correlates of the second factor and its item content reflect more than impulse suppression alone, although certainly that is the factor's core. The data suggest that the *A* and *R* scales tap a cluster of characteristics one would expect in individuals who are prone to anxiety and who chronically use repressive defenses. Both scales, though, are to be understood in a characterological, trait sense rather than in a process sense.

The means and standard deviations for both samples are given in Table 1. There are no significant differences in mean *A* or *R* scores between samples for either sex, nor do the sexes differ in mean *R* score. There is, however, a significant difference ($p < .01$) between the sexes in mean *A* score.

Incidental Cue-Utilization Procedure

The procedure for assessing the use of incidental cues in problem solving has been described in detail in previous papers by the present writers (Mendelsohn & Griswold, 1964, 1966). Consequently, only a brief outline will be presented here.

After the administration of the personality inventory, subjects were given a list of 25 five-letter words to memorize (focal stimuli), while another list of 25 words were being played on a tape recorder in the same room at ordinary speaking volume (peripheral stimuli). They were told that this part of the experi-

TABLE 1

MEANS AND STANDARD DEVIATIONS OF *A* AND *R* SCORES

Sample	*A*				*R*			
	Males		Females		Males		Females	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	11.00	6.70	15.24	6.62	18.30	5.38	17.83	4.12
2	10.88	6.93	13.27	6.56	18.98	4.30	18.30	4.39

---

[2] The Luborsky et al. (1965) study provides support for this assumption, but the present study was conducted prior to the appearance of their paper.

ment was a study of "memory under interference conditions," a set which led them, in general, to attempt to screen out the peripheral stimuli. Subjects were also told that they would have 10 minutes to memorize the list, after which they would be asked to solve some problems and then to recall the list. When the 10 minutes were up, subjects were given a set of 30 single-solution, five-letter anagrams to solve. Each anagram was projected on a screen by a slide projector for 10 seconds and subjects were asked to write down only the solution word. Ten of the solution words had previously appeared on the focal (memory) list and 10 on the peripheral (interference) list[3]; the remaining 10 had not previously appeared at all in the experimental situation and served as an estimate of anagram-solving ability independent of the influence of the cues.

After the presentation of the anagrams, the subjects were given 5 minutes to recall the focal list and then another 5 minutes to recall as much as they could of the peripheral list. In obtaining recall measures, only those 15 words of each list which were not anagram solutions were used. This seemed the preferred measure of recall because (*a*) at the time of exposure, subjects could not distinguish between these "filler" words and the cue words, and (*b*) they had no subsequent experience with them, as they did with the cue words, during the experiment.

RESULTS

In order to control for the effect on focal and peripheral solutions of anagram-solving ability per se, the data were analyzed by means of analyses of covariance in which number of neutral solutions served as the covariate. It will be shown shortly that neither *A* nor *R* scale scores are significantly related to the solution of the neutral (uncued) anagrams. Thus, the effect of the covariance procedure is to reduce error substantially, but to alter the main and interaction effect sums of squares very little.

Relationship of R to Cue Utilization

Since the means and standard deviations of the *R* scale scores are quite comparable for the male and female subjects, their data were combined in a single analysis. The main effects were level of *R* (low, 16 and below; middle, 17–20; high, 21 and above), sex of subject, sample, and type of anagram (focally or peripherally cued). The last named is a within-subjects effect. In order to obtain equal numbers of males and females and proportionality of cell *n*'s, 43 subjects (37 of them females) were dropped at random from the sample. There were thus 23 males and 23 females at each level of *R*. The results of the analysis of covariance are shown in Table 2. The means for neutral, focal, and peripheral solutions are shown in Part I of Table 3 and the adjusted focal and peripheral solution means in Part II of the Table.

TABLE 2

ANALYSIS OF COVARIANCE FOR *R*: BOTH SEXES AND SAMPLES

Source	*df*	*MS*	*F*
Between *S*s			
R	2	26.42	7.89***
Sample (B)	1	3.36	1.00
Sex (C)	1	.73	<1
R × B	2	5.85	1.75
R × C	2	3.52	1.05
B × C	1	7.14	2.13
R × B × C	2	3.50	1.04
Error between	125	3.35	
Within			
Focal-Peripheral (D)	1	196.70	76.24***
R × D	2	1.90	<1
B × D	1	10.20	3.95*
C × D	1	2.64	1.02
R × B × D	2	.16	<1
R × C × D	2	4.34	1.68
B × C × D	1	.03	<1
R × B × C × D	2	4.15	1.61
Error within	126	2.58	

* $p = .05$.
*** $p < .001$.

The major hypothesis that *R* scores would be negatively related to the use of incidental cues is clearly confirmed by the data; with anagram-solving ability controlled, the low-*R* subjects solve more of the cued anagrams than do the

TABLE 3

SOLUTION MEANS FOR NEUTRAL, FOCAL, AND PERIPHERAL ANAGRAMS

Anagram	Part I: Unadjusted means					
	Males			Females		
	Low	Middle	High	Low	Middle	High
Neutral	3.35	3.65	3.52	3.91	4.00	3.52
Focal	5.74	5.00	4.57	6.17	5.22	4.61
Peripheral	4.22	3.78	2.83	3.70	3.17	3.48
	Part II: Adjusted means					
	Males			Females		
	Low	Middle	High	Low	Middle	High
Focal	5.88	5.00	4.63	6.06	5.06	4.67
Peripheral	*4.36*	*3.78*	*2.89*	*3.59*	*3.01*	*3.54*
Sum	10.24	8.78	7.52	9.65	8.07	8.21

---

[3] Different anagram solutions were cued for the two samples.

middle- or high-R subjects. Reference to the first part of Table 3 makes this finding even more striking, for it is apparent that the high-, middle- and low-R subjects do not differ from each other in number of neutral (uncued) solutions ($F < 1$); in fact, the group which had the lowest neutral solution mean, the low males, solved more of the cued anagrams than any other group. Even without covarying neutral solutions, the levels of R main effect for focal and peripheral solutions yields an F ratio of 5.59, significant beyond the .01 level.

The absence of any significant interactions between R and sex, sample, or type of anagram indicates that the basic finding is replicated, holds for both sexes and for both focally and peripherally cued anagrams. Although none of the interaction effects approaches significance, it can be seen in Table 3 that the results are more emphatic for males than for females and, in particular, that there is little evidence that R is related to the differential use of peripheral cues by females. It can also be seen in Table 3 that the number of focal solutions is considerably greater than the number of peripheral solutions, a result which, given the manner in which the two sets of cues are presented, is to be expected.

In their paper on leveling and repression, Holzman and Gardner (1959) reported that although extreme female repressers were levelers, extreme levelers were not necessarily repressers. Following this lead, the present data were examined further to ascertain whether or not the relationship of repression and cue utilization is reciprocal. Subjects were divided into high-, middle-, and low-cue utilizers on the basis of their summed adjusted focal and peripheral scores and the mean R scores for each group examined. The results indicate that for males the relationship is reciprocal—the high-cue utilization group has a lower mean R score than the middles, who in turn are lower than the low group ($\bar{X}$ high $= 16.83$, middle $= 18.88$, and low $= 20.33$, $p < .05$). In contrast, the differences among the female groups are scarcely discernible. The results for the females, then, parallel those obtained by Holzman and Gardner.

Relationship between A *and Cue Utilization*

Since the mean A scores are significantly different for males and females, separate analyses were run for each sex. Again, subjects were dropped at random from the sample in order to obtain proportionality of cell n's.

Females. The three female groups were defined as follows: low, 10 and below; middle, 11–16; high, 17 and above. The analysis of covariance produced no significant main or interaction effects associated with level of A. Nevertheless, a further analysis of variance, contrasting the mean A scores of high, middle and low groups of cue utilizers was run, but again no significant results emerged.

TABLE 4

ANALYSIS OF COVARIANCE FOR A: MALES IN BOTH SAMPLES

Source	*df*	*MS*	*F*
Between *S*s			
A	2	17.55	5.30**
Sample (B)	1	4.54	1.37
A × B	2	1.81	<1
Error between	59	3.31	
Within *S*s			
Focal-Peripheral (C)	1	66.94	26.25***
A × C	2	2.60	1.02
B × C	1	3.51	1.38
A × B × C	2	.18	<1
Error within	60	2.55	

** $p < .01$.
*** $p < .001$.

Males. The three male groups were defined as follows: low, 7 and below; middle, 8–12; high, 13 and above. As with R, there are no significant differences among the groups in number of neutral solutions ($F < 1$), the means being low, 3.14; middle, 3.41; and high, 3.55. Thus, the adjusted means shown in Table 5 differ only slightly from the unadjusted means. The results of the analysis of covariance are shown in Table 4. Although the main effect for A is significant at beyond the .01 level, inspection of the adjusted means in Table 5 indicates that, contrary to prediction, the relationship of A to cue utilization is curvilinear. This unexpected result is the more striking because it holds in remarkably similar fashion for both sets of anagrams, focal and peripheral, and in both samples; in all cases, the middle-A group does least well.

As one would expect from these results, there are no significant differences in mean A score

TABLE 5

ANALYSIS OF COVARIANCE FOR A: ADJUSTED MEANS, MALES

Anagram	Sample 1			Sample 2		
	Low	Middle	High	Low	Middle	High
Focal	4.84	3.97	5.08	5.19	4.48	6.34
Peripheral	*4.04*	*3.07*	*3.58*	*4.02*	*2.81*	*4.01*
Sum	8.88	7.04	8.66	9.21	7.29	10.35

among the high-, middle-, and low-cue utilization groups.

The final step was to examine the data for the possibility of interactions between A and R. The separate analyses of the male and female samples produced no significant replicated interactions.

Memory

Recall measures were obtained for both the focal (memorized) and peripheral (interference) lists after the anagram-solving task. These data were used in two ways: first, to evaluate the hypothesis that the differential use of the cues in anagram solution is attributable to differential recall of the cues, and second, to examine relationships between A and R and recall. For the reasons given previously, only the "filler" words were used to derive recall measures.

The correlation of the focal recall scores with the number of focal solutions (adjusted for neutral solutions) is only .078, and the equivalent correlation for adjusted peripheral scores is −.003. Neither correlation is of sufficient magnitude to support an interpretation of the obtained results in terms of the effects of differential recall. Despite this result, a moderate correlation between R scale scores and recall of peripheral words was obtained in both samples for male subjects. In Sample 1, $r = -.30$, in Sample 2, $r = -.35$, and for the samples combined, $r = -.32$, $p < .01$. In contrast, the correlation for females is a negligible .05. The correlations of R and focal recall scores are insignificant for both sexes in both samples. Likewise, no significant linear or curvilinear relationships between A scores and recall measures were found.

These findings can be summarized as follows:

1. As predicted, there is a negative relationship between scores on the R scale and the degree to which incidental cues are utilized in anagram solution.

2. This relationship holds for incidental cues appearing in both the focus and in the periphery of attention for males, but only for those appearing in the focus of attention for females.

3. The relationship between R scale scores and cue utilization is reciprocal for males in that high-R subjects utilize the cues least, and low-cue utilizers are high-R subjects, etc., but this reciprocality does not obtain for females.

4. Contrary to prediction, scores on the A scale are not related in a negative and linear fashion to the use of incidental cues in problem solving. However, for males only, A is related in a curvilinear fashion to the utilization of both focal and peripheral cues, the middle-A group using the cues least.

5. Recall of the incidental cues is not significantly related to the measures of their use in problem solving. There is, for males only, though, a significant negative relationship between R scores and recall of the cue words presented in the periphery of attention.

All the results summarized above were replicated in the two samples used in the experiment.

DISCUSSION

Any attempt to interpret the present results depends in part on an understanding of the psychological processes involved in the cue utilization procedure. In previous papers (Mendelsohn & Griswold, 1964, 1966) we have argued that the observed enhancement of performance is dependent upon a wide deployment of attention by subjects; that is, in order for subjects to use the cues they must first attend to them in some manner. Since both sets of cues are presented at the same time, some subjects may attend to the focal list and screen out the peripheral list, some may find their attention drawn to the peripheral list to the detriment of the focal list, some may be merely distracted by the peripheral list and thus find it difficult to attend to either list, and others may be able to attend to both lists in a maximally effective way. If this analysis is correct, those subjects who show the greater overall enhancement of performance are those who are able to deploy attention most widely and effectively. It should be noted that "to deploy attention most widely and effectively" does not mean broad scanning alone, but also the ability to retain that which has been scanned.[4]

Given this interpretation, the results would seem to indicate that subjects who chronically use repressive defenses have a relatively narrow range of attention deployment. This finding is consistent with Schafer's (1948) formulation on the relationship of repression and attention in test performance, and with the results of the Luborsky et al. (1965) study. The present finding is of particular interest, however, because the stimuli used had no apparent conflict relevance. This suggests that there is a characteristic perceptual approach (or strategy) used by repressers which is not limited to situations in which threat or impulse is present, although it may be intensified in such situations. The results of Luborsky

[4] Evidence in support of this interpretation is given in Mendelsohn and Griswold (1966). The question of the way in which the cues are retained is also considered in that paper.

et al., for example, are most clear-cut for sexual stimuli.

Although the present data are limited, it is reasonable to inquire into the adaptive function of such a characteristic narrowing of attention. Repressive defenses are not absolutely effective in their operation-breakthroughs of unconscious material; impulses or painful self-knowledge can and do occur. Psychoanalysis with its typical emphasis on intrapsychic events interprets such breakthroughs primarily in terms of increases in the level of drive. Another (though not contradictory) possibility is that repressed material can be elicited by stimuli appearing in the external environment. If a stimulus of unmistakable relevance to repressed material comes into attention, it may be impossible to maintain the defense. This line of reasoning is illustrated by a case seen by the first writer, in which a young man was overwhelmed by the most primitive and undisguised, though previously repressed, Oedipal and pre-Oedipal fantasies after he had unexpectedly seen his parents having intercourse. While this is an extreme case, the point is that given the stimulus, repression was no longer a possibility.

From this standpoint, the narrowing of the range of attention may be seen as an attempt to protect the effectiveness of the repression. This view assumes that it is easier to control attention to focal than to peripheral stimuli—focal attention can be directed to what is safe or away from what is potentially threatening. However, there is less voluntary control over the reception of peripheral stimuli; they can impinge on awareness without the intention of the perceiver. Perhaps the most effective way to deal with the possibility of threatening stimuli appearing in this unexpected and uncontrollable fashion is to narrow the range of attention and screen out all but that which one intends to perceive. Repression in this view consists of two processes (which may develop sequentially or simultaneously), an inhibition and a perceptual strategy which protects and may eventually obviate the inhibition.

No differentiation has been made between males and females in this discussion. In one sense this is justified, since no significant interactions between level of *R* and sex were found. In another sense, however, distinctions should be made, for the results are stronger for males than for females; there is no relationship between *R* and recall of the incidental list for females, and the relationship between the use of cues and *R* is not reciprocal for females as it is for males. That some differences between the sexes were found is not entirely unexpected since Holzman and Gardner (1959) obtained a significant relationship between repression and leveling for only one sex, in their case, females. In the absence of appropriate data within the present experiment or of any theorizing about defense which leads to different expectations for males and females, one can only discuss possibilities. Three seem most likely: (*a*) *R* scale scores have different meanings for males and females; (*b*) the cue utilization procedure would be understood differently for males and females; or (*c*) the operation of defenses is different for males and females. The first possibility receives independent empirical support in Block (1965), who concluded that in measuring the second dimension of the MMPI, ". . . here is one instance, we believe, where a substantial improvement in the measurement of a dimension may result if separate scale construction [for the sexes] is applied [p. 116]." The findings of Holzman and Gardner give some independent support to the third possibility too. This seems a fruitful area for future research, but for the present it should be noted that the sex differences are of degree and not of direction. It seems reasonable, then, to speak of the relationship between cue utilization and repressive defenses in a general, but cautious, way. The necessity of taking sex differences into account, however, should be evident.

One of the most intriguing aspects of the work on the relationship of "cognitive style" to defense is the nature of that relationship, that is, does one grow out of the other or do they develop simultaneously, etc.? A good discussion of the problem can be found in Holzman and Gardner (1959). In their paper they tend to favor the position that, because repressers are levelers but levelers are not necessarily repressers, "Repression relies in part, but by no means exclusively, upon the process of leveling for its execution [p. 154]." The lack of a reciprocal relationship between *R* and cue utilization, similar in form to that between repression and leveling, may point to a parallel conclusion for females in the present study. However, the presence of a reciprocal relationship for males makes more tenable the view that the "cognitive style" and the defense are aspects of the same underlying mechanism or behavioral orientation. We may have here a reflection of an important difference between the defensive operations of males and females.

The hypothesis that *A* would be related in a negative and linear fashion to cue utilization was clearly not confirmed. The most direct explanation of the failure is that previous experimenters and reviewers have, on the whole, been concerned with *states* of emotion or anxiety rather than with the trait of proneness to anxiety or chronic anxiety. However, two points militate against easy acceptance of this view; first, some studies using trait-anxiety measures have shown a reduced range of attention (most relevantly, Silverman & Blitz, 1956), and in the present study *A* scores are related in a consistent, though not linear, fashion to the use of cues by males. An experiment by Cornsweet (1964) provides a possible explanation. She found that when peripheral cues are relevant to the central task, shock threat increases the extent to which they are used. In most experiments on arousal, anxiety, and cue utilization, the peripheral or incidental stimuli have had no relationship at all to the central task. In the present experiment, the incidental cues are clearly relevant to the anagram-solving task. Perhaps the comparative enhancement of performance observed in the high-*A* males can be understood in terms of the Cornsweet finding, while that of the low-*A* males is attributable to an habitually wider range of attention deployment. Admittedly, this is a speculative explanation, but it is based on the assumption that while Easterbrook's (1959) conclusion that emotion reduces the range of cue utilization may hold generally, a corollary statement is needed: within limits, arousal increases attention to those cues, focal, peripheral or incidental, which are relevant to adaptive performance. Once again, neither the present data nor that of others permits an explanation of the obtained sex differences.

In general, this experiment supports the idea that anxiety and repression, measured as traits, have important concommitants in characteristic perceptual behavior. However, the results make it clear, too, that the assumption that anxiety and repression have identical correlates and effects for males and females is of doubtful validity. This latter point needs to be stressed, for research and theorizing in this general area have more often than not ignored the necessity of taking sex differences into account.

REFERENCES

Block, J. *The challenge of response sets.* New York: Appleton-Century Crofts, 1965.

Byrne, D. Repression-sensitization as a dimension of personality. In B. A. Maher (Ed.), *Progress in experimental personality research.* Vol. 1. New York: Academic Press, 1964. Pp. 170-220.

Cornsweet, D. Use of cues in the visual periphery under conditions of arousal. Unpublished doctoral dissertation, University of California, Berkeley, 1964.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* 1959, **66**, 183-201.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H., & Spence, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues,* 1959, **1**, 1-186.

Holzman, P. S., & Gardner, R. W. Leveling and repression. *Journal of Abnormal and Social Psychology,* 1959, **59**, 151-155.

Korchin, S. J. Anxiety and cognition. In C. Scheerer (Ed.), *Cognition: Theory, research, promise.* New York: Harper & Row, 1964. Pp. 58-78.

Luborsky, L., Blinder, B., & Schimek, J. Looking, recalling and GSR as a function of defense. *Journal of Abnormal Psychology,* 1965, **70**, 270-280.

Mendelsohn, G. A., & Griswold, B. B. Differential use of incidental stimuli in problem solving as a function of creativity. *Journal of Abnormal and Social Psychology,* 1964, **68**, 431-436.

Mendelsohn, G. A., & Griswold, B. B. Assessed creative potential, vocabulary level, and sex as predictors of the use of incidental cues in verbal problem solving. *Journal of Personality and Social Psychology,* 1966, **4**, 423-431.

Schafer, R. *The clinical application of psychological tests.* New York: International Universities Press, 1948.

Schafer, R. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954.

Silverman, R. E., & Blitz, B. Learning and two kinds of anxiety. *Journal of Abnormal and Social Psychology,* 1956, **52**, 301-303.

Welsh, G. S. Factor dimensions *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: University of Minnesota Press, 1956. Pp. 264-281.

(Received July 6, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 360–364

COMPLIANCE, EXPECTATION OF REWARD, AND OPINION CHANGE[1]

HAROLD B. GERARD

University of California, Riverside

The effect of expectation of reward for forced compliance was examined experimentally. In a factorial design under 1 treatment S received less, more, or the amount of money he expected for taking a discrepant stand. In another treatment he received 1 of the 3 reward levels without any prior expectation. The data indicate a curvilinear trend with greatest postadvocacy change at the moderate reward level. In general, the data can best be interpreted as having been due to an incentive effect. There was some support, however, for dissonance theory in that at the lowest reward level there was more opinion change in the direction of the position advocated under the expectation treatment where it was more clearly an insufficient reward.

An issue about which there is currently a good deal of controversy concerns the relationship between the amount of reward offered a person for advocating an opinion that is discrepant from his private opinion and the amount of his subsequent opinion change in the direction of the opinion he advocated. In the experiment described here the role of expectation of reward as it affects opinion change under these conditions will be examined. The person's expectation is a feature of the situation that is acknowledged as being important but one that has as yet not been examined systematically.

The experiment by Festinger and Carlsmith (1959) which started the controversy was based upon a derivation from dissonance theory that predicts an inverse relationship between the amount of reward offered to a person for publicly advocating an opinion that is discrepant from his actual opinion and the subsequent change in his actual opinion in the direction of the opinion he advocated. This relationship appears, at first glance, to be paradoxical; the less justification the person has for advocating the discrepant opinion, the greater will be the amount he changes his opinion toward the stand advocated and hence away from his own private conviction. There is no paradox, however, when we consider that opinion change is itself an after-the-fact way of justifying having taken the discrepant stand in the first place. After deciding to advocate an opinion he initially does not agree with, the person comes to find this opinion more acceptable and thus provides justification for his decision. The theory entertains a variety of ways in which the person may reduce dissonance in this type of situation other than opinion change, but we shall not consider these other avenues here. Festinger and Carlsmith confirmed the prediction within a context where a subject, after participating in a very boring task, was paid either \$1 or \$20 to lie to another subject by telling him that the task was interesting. A postmeasure of the subject's attitude toward the boring task indicated that he found it less boring in the \$1 than in the \$20 condition. This was offered as evidence for the theory. In the \$1 condition the subject had less justification for lying than he did in the \$20 condition and hence experienced greater dissonance which he subsequently reduced by justifying his lie; he came to believe that the task was indeed not as boring as he originally thought it was.

A number of criticisms were raised regarding this interpretation of the results, one of these being that the effect was not due to the need for justification in the \$1 treatment but to possible suspicion aroused in the \$20 treatment. The payment, that is, was far beyond the subject's *expectation*. The argument is that a subject would be suspicious about being offered \$20 to tell such a harmless lie and would therefore be negatively disposed toward the experiment in general, resulting in less of a tendency for him to change in the direction of the opinion he was asked to advocate.

In a subsequent experiment Cohen (reported in Brehm & Cohen, 1962) attempted to answer this criticism by introducing four levels of reward: \$.50, \$1, \$5, and \$10. His subjects were college students, and the issue they chose was a very salient "town" versus "gown" dispute, concerning which the students were unanimously (or nearly so) on the gown side. Each subject was asked to write an essay favoring the town position on the issue. A measure of the subjects' subsequent attitudes toward the issue showed an inverse relationship between amount of payment and favorableness toward the town side of the argument, thus adding strong confirmation to the findings in the Festinger and Carlsmith experiment and effectively putting to rest the "suspicion" interpretation of the original results.

---

[1] This study was conducted with support from Grant MH 118170 from the National Institute of Mental Health and Grant GS 392 from the National Science Foundation. I would like to thank Edward F. O'Connor, Jr. for serving as the experimenter and helping with the data analysis.

Two recent experiments, one modeled after the Festinger and Carlsmith procedure and the other after the one by Cohen, have reopened the issue as to whether a dissonance interpretation of the earlier experiments is appropriate. Janis and Gilmore (1965) resurrected the "suspicion" interpretation of the Festinger and Carlsmith results. The subject in their experiment was asked to write an essay in favor of the proposition that a year of physics and a year of mathematics should be required of all college students. As in the Festinger and Carlsmith experiment, the subject was offered either $1 or $20 if he agreed to write the essay. In the Festinger and Carlsmith experiment the subject agreed not only to lie to the next subject but also to be on call to lie to other subjects in the future. A $20 payment for all these tasks would not tend to arouse as much suspicion as the procedure used by Janis and Gilmore of merely writing an essay. Subjects were indeed found to be more suspicious in their $20 treatment than in the one used by Festinger and Carlsmith. They might have contrasted a condition in which $20 was *clearly* a justified eward with the procedure used by Festinger and Carlsmith, where there is certainly some doubt bout how justified that large a reward might ave seemed to the subject. They found a staistically insignificant trend in the opposite direcion; that is, there was a slightly greater tendncy for the subject in the $20 treatment to end p more favorably disposed toward adding scince and math to the curriculum than in the $1 eatment. This is certainly not evidence for the uspicion hypothesis which Janis and Gilmore resumably started out to test. It is also probay true that their manipulation did not provide n adequate test of the hypothesis.

Nowhere in their report do Janis and Gilmore ention what their subjects' private convictions ere initially. The inverse relationship between titude change and reward predicted by disso-nce theory requires that the subject take a *screpant* stand. If he writes an essay that rees with his conviction, we would, on the basis an incentive effect, expect a positive relationip between attitude change and reward. There no way of knowing, at least from the report of the procedure given by Janis and Gilmore, how many subjects were pro science and math and how many were anti to start with. Their procedure assumes that all subjects were initially anti science and math. From the standpoint both of a failure to test the suspicion hypothesis and the possibility that a number of subjects were not advocating a discrepant stand, the Janis and Gilmore results must be considered inconclusive.

Rosenberg (1965) has offered another interpretation of the inverse relationship between amount of reward and counterattitudinal opinion change. He argues that since the typical subject is probably concerned with the experimenter's evaluation of him he would not like to be judged as immature, maladjusted, or dishonest. In an experimental treatment where the subject is offered a large reward for engaging in certain behavior, ". . . he may be led to hypothesize that the experimental situation is one in which his autonomy, his honesty, his resoluteness in resisting a special kind of bribe, are being tested [p. 29]." To the extent that he entertains this hypothesis, the subject will be likely to show stalwart resistance to exhibiting the behavior he perceives the bribe was intended to elicit. In the case of forced-compliance studies he may perceive that in addition to advocating the discrepant stand for the reward offered, the experimenter is bribing him to change his attitude. It follows from this argument that the larger the reward, the more resistant the subject will be to indicating that he had changed his attitude.

In testing this hypothesis, Rosenberg repeated the Cohen experiment using a different issue but one on which student opinion was polarized. A significant difference between the two procedures was that in the Cohen study the request to write the essay was made by the same experimenter who took the final attitude measurement, whereas in the Rosenberg experiment the request was made by one experimenter and the measure taken by another. This should make no difference as far as the dissonance-theory prediction is concerned, but should make a difference as far as "evaluation apprehension" is concerned. Rosenberg found, using rewards of $.50, $1, and $5, a *positive* relationship between attitude change and the size of reward, which supports his interpretation.

Rosenberg did not run a set of conditions paralleling his own which exactly replicated Cohen's procedure. Since he did use a different subject population, these conditions would have provided an important control. Also his notion about evaluation apprehension does not explain the Festinger and Carlsmith results, since in that experiment the postmeasure of opinion *was* taken by someone other than the experimenter who offered the reward.

One must also raise questions about specific aspects of Rosenberg's procedure that differed from the procedure used by Cohen, and that might therefore account for the results which were opposite to those of Cohen. One such difference between the Cohen and Rosenberg experiments, which may be a key to the discrepancy in results, is the difference in degree of affluence of the two subject populations used. Cohen's experiment was conducted with Ivy League students, whereas Rosenberg used students at a state university. Perhaps for a state university student $.50 would seem a perfectly adequate payment for writing a brief essay, $1 would seem generous, and $5 more than generous. A typical Ivy League student, on the other hand, might consider anything less than $10 an insufficient reward for the effort required of him. If there were such a difference in the perceived value of money, it would account for the difference in results. Dissonance is aroused only when the reward for compliance is perceived as *insufficient,* a condition which may not have been satisfied in the Rosenberg experiment where the results might have been due to an incentive effect, since a reward greater than $.50 may have provided further inducement. When replicating an experiment, the important thing is not necessarily to repeat the experiment in exact detail, but to create the conditions required to test the hypothesis even if it may mean altering certain aspects of the procedure in order to do so. In the case of this series of experiments, a proper test of dissonance theory ultimately hinges on whether or not sufficient and insufficient reward conditions are being contrasted. With a change in subject population, rewards may have to be altered or certain other techniques employed in order to produce the psychological states required to test the hypothesis. Cohen may have made a felicitous choice of reward levels and Rosenberg may not have. Lacking further information as to the perceived sufficiency or insufficiency of the rewards in both experiments, this interpretation of the difference in results can only be a tentative guess.

Since the perception of insufficiency of reward is crucial, the subject's expectation as to what is a sufficient reward must somehow be violated. Offering him a smaller reward than he expected would produce dissonance and a subsequent tendency to reduce this dissonance by attempting to justify having engaged in the discrepant behavior. On the other hand, offering him a more than sufficient reward would tend to produce an incentive effect.

METHOD

An Overview of the Design

Two experimental treatments were compared, one in which the subject was led to expect a $2 honorarium for participating in a "survey of student opinion" on a current campus issue, and one in which no expectation of a sufficient reward was induced. Within each of these treatments, the subject was given either $.50, $2, or $5 to write an essay advocating the side of the issue he disagreed with. The subject's subsequent opinion on the issue was measured to determine the amount of his opinion change away from his original stand on the issue.

Procedure

The subjects were 48 male undergraduates at the University of California, Riverside, who were approached individually in their dormitory rooms and assigned randomly to the experimental conditions. The experimenter introduced himself as a fellow student who was working for the Survey Research Institute of Los Angeles, a (fictitious) research organization which was doing a study of various aspects of the free speech movement which had been active at the University during the school year. If the subject agreed to let the experimenter into his room, he was handed a sheet of paper briefly describing the purpose of the survey and indicating that the topic of particular interest at the moment was the controversial issue of whether or not the student government should be able to act or take positions on non-University-related issues. The regulations that were in force prohibited such action. The Survey Research Institute, he was told, was involved in finding out the pros and cons on the issue.

In the expectation treatment there was an additional paragraph to the effect that each student participating in the survey would receive a $2 honorarium. This paragraph was omitted in the no expectation treatment.

In the expectation treatment, the subject in the $.50 condition was handed $.50 and was told, "unfortunately, we have had some unanticipated expenses and we can only afford to pay $.50 instead of the $2 that we had planned to give each student." In the $2 treatment the experimenter handed the subject $2. In the $5 condition he was handed $5 and was told, "fortunately we have had fewer expenses than we had anticipated and we can afford to pay $5 rather than $2 to all the students who help us in this survey. Since we would have had money left over, we thought this was an equitable way of distributing that money." In the no-expectation treatment the subject was given either $.50, $2, or $5 and was told that he was being given the money in return for his help with the survey.

At this point the subject was handed a sheet of paper on which to indicate whether or not he was in favor of the present regulation that forbids the student government to take a position on off-campus issues. After having done this, he was asked to write an essay supporting the side of the issue he opposed. The excuse given for asking him to do this was that the Institute had a sufficiently large number of essays that favored the position he believed in, and arguments were needed on the other side. These essays were to be seen by the staff of the Institute. Only one subject refused to comply with this request. He was replaced in the design. After the subject completed his essay, he stated his present opinion on the issue by checking a point on a scale 15 centimeters long. The true purpose of the experiment was then revealed, and the subject was asked to keep this information in strictest confidence. Very little difficulty was experienced in asking subjects in the $5 treatment to return the money. Each subject was given $2 for participating in the experiment.

Summary of the Design

The experiment used a 2 × 3 factorial design; three levels of monetary reward were run under a treatment where a prior expectation was induced or under a treatment where there was no prior expectation. In the expectation condition the reward was less than, equal to, or more than the subject expected to receive. This was done in an attempt to induce psychological states of insufficient, sufficient, or oversufficient reward. Eight subjects were assigned to each treatment.

RESULTS

The data from the experiment consisted of the subject's attitude position on the issue as measured by a postexperimental questionnaire. He checked a point on the 15-centimeter bipolar scale indicating the extent of his agreement or disagreement with the present campus regulation. One end of the scale was labeled "the regulation is completely justified" and the other end was labeled "the regulation is completely unjustified." It was assumed that the subject's initial position on the issue was at either one of the two extremes, depending upon how he responded to the premeasure. The distance from this extreme to the point he checked on the scale was taken as the change measure; the larger this distance, the greater the assumed change. For the purposes of measurement the scale was divided into 24 equal units.

The data, which are presented in Table 1, show a curvilinear relationship between attitude change and reward under both expectation and no expectation. The difference between expectation and no expectation is not significant. There is, however, an overall effect of reward level ($p < .05$, F test). Further analysis indicates a significant difference between the $2 and $.50 treatments ($p = .01$, t test). The difference between the $2 and $5 treatments is not statistically significant.

TABLE 1

EFFECT OF EXPECTATION AND REWARD ON OPINION CHANGE[a]

Reward level	Expectation	No expectation
$ .50	4.56	1.62
$2	7.25	6.81
$5	5.19	5.12

[a] The larger the number, the greater the change away from the subject's original opinion.

DISCUSSION

If we assume that there is no reliable difference between the $2 and $5 treatments, and this is probably a safe assumption to make since the edge that appears to favor the $2 treatment is due primarily to one large score under each of the two expectation conditions, the data neither confirm dissonance theory nor do they show a clear effect of incentive. For dissonance theory to have been confirmed there should have been an inverse relationship between opinion change and reward under the expectation treatment. Assuming that our subjects and those of Rosenberg enjoy the same degree of affluence, we would have expected results similar to Rosenberg's under the no expectation treatment, although we had no clear prediction of the direction of the relationship. If the direction of the relationship between opinion change and reward had been positive under no expectation and negative under expectation, our intuition as to the discrepancy between the Cohen and Rosenberg experiments would have been confirmed, namely, that the difference was due to a difference in affluence of the two subject populations.

The results do suggest the operation of incentive if the $.50 and $2 treatments are compared. For an incentive theory to have been unequivocally supported, however, the $5 treatment should have produced more opinion change than the $2 treatment. There was nothing like a "ceiling effect" since the subjects in the $5 treatment could have changed considerably more than they did.

When the two $.50 conditions are compared, there is a suggestion that dissonance was operating since the change in the expectation condition was greater than in the no-expectation condition ($p = .05$, t test). There is amazingly little overlap (only 4 cases out of 16) between the two $.50 conditions ($p < .05$, chi-square test).

This interpretation assumes that a feeling of insufficient justification was induced in the $.50-expectation condition.

Thus, we find partial evidence for the operation of both dissonance and incentive. Part of the problem may be due to the unexpected importance of one aspect of the procedure. You will recall that when the subject was approached in his dormitory room the experimenter entered the room, introduced himself, and stated his purpose. Once the subject had invited the experimenter into his room this was tantamount to complying with nearly any reasonable request the experimenter might make that pertained to the opinion survey. The subject at that time did not expect to receive any monetary reward. When the reward was offered it may have served as an additional incentive. This is likely to have happened in the no-expectation treatment and possibly in the expectation treatment as well. Thus, we may not have produced the basic conditions necessary for testing a dissonance versus incentive interpretation of the earlier experiments since it may be necessary to get the subject to commit himself for what he believes to be an insufficient reward *before* he permits the experimenter to enter his room. It is the *decision* to comply for an insufficient reward that presumably produces the dissonance.

Upon examining the Cohen and Rosenberg procedures we find that they did differ in this respect. Cohen used the monetary reward in order to get the subject to commit himself to writing the essay, whereas Rosenberg followed the procedure used in the present experiment by having the subject essentially commit himself to cooperating with the experimenter and then offering him the monetary reward. The fact that our results do tend to resemble Rosenberg's rather than Cohen's suggests that this may be a crucial aspect of the procedure and one that should be given special attention in future research in this area. The difference we found between the two $.50 conditions suggests that in spite of the possible flaw in our procedure we did induce a psychological state of insufficient reward under the expectation treatment.

A procedure that might be used to realize a more exacting test of the two interpretations would be to somehow create an expectation of a specific monetary reward in one treatment and not in the other. Then, with only one foot barely in the door of the subject's dormitory room, offer him one of three reward levels for writing a discrepant essay—one less than, one equal to, and one greater than the reward promised in the expectation treatment.

REFERENCES

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance*. New York: Wiley, 1962.

FESTINGER, L., & CARLSMITH, J. M. Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology,* 1959, 58, 203–210.

JANIS, I. L., & GILMORE, J. B. The influence of incentive conditions on the success of role playing in modifying attitudes. *Journal of Personality and Social Psychology,* 1965, 1, 17–27.

ROSENBERG, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology,* 1965, 1, 28–42.

(Received July 6, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 3, 365-370

AWARENESS, LEARNING, AND THE BENEFICENT SUBJECT AS EXPERT WITNESS[1]

LEON H. LEVY

Indiana University

In order to determine the extent to which inferences of awareness in verbal conditioning may be validly derived from Ss' replies in postconditioning awareness interviews, 16 Ss were given full information about the experiment by a confederate posing as another S before being run in a Taffel verbal conditioning procedure and compared with 16 uninformed Ss, with respect to their learning and interview behavior. The 2 groups did not differ on rate of learning, but did on level of performance, with the informed group being superior. In the interview, a significantly larger number of questions were required to elicit reports of awareness from informed than from uninformed Ss, and only 25% of the informed Ss admitted to any form of prior knowledge. The implications of these findings both for the assessment of awareness and for the interpretation of the processes involved in verbal conditioning were discussed.

There are certain issues in psychology for which it appears that even the most rigorous and behavioristically inclined researchers are apt to turn to their experimental subjects as expert witnesses. One such issue is the role of awareness in human learning (Adams, 1957). Beginning with Thorndike's (1932) assertion that learning can occur without awareness, and continuing through to the current controversy over the role of awareness and cognition in verbal conditioning (Spielberger, 1965), psychologists have based their positions in large part upon the verbal testimony of their subjects. Those who would dismiss awareness as a scientifically useful concept (Verplanck, 1962), no less than those who would embrace it (Dulany, 1962; Spielberger, 1962), have turned to their subjects' verbal reports in support of their positions. Farber (1963), for example, has said that awareness cannot be disregarded in the study of behavior and that one excellent approach is to "ask the subjects what they are doing or think they are doing and why [p. 195]." As straightforward as this approach might be, it is obviously not without its problems.

It seems to be generally accepted that awareness is a private event of which only the subject has knowledge: while the experimenter may gather data on learning via response frequencies, latencies, and the like, only the subject has access to data on awareness. Thus, to the extent that his verbal reports of awareness are taken as veridical or given the same weight as the experimenter's summary statistics, the subject may be said to be cast in the role of co-investigator or expert witness (Buck, 1961). This is a role which he has not formally played since the days of introspectionist psychology, and because so many crucial theoretical issues have been made to rest upon his testimony, it becomes a proper subject of investigation.

The present study was conducted within a verbal conditioning framework for both theoretical and methodological reasons. Verbal conditioning has come to occupy a focal position in theoretical controversies over the nature of complex human learning and psychotherapy (Krasner, 1958, 1962; Spielberger & Levin, 1962), and in the test of a variety of hypotheses in personality and psychotherapy (Williams, 1964). To a large degree, the issues hinge upon whether the learning curves are interpreted as manifesting the unconscious and automatic effects of reinforcement upon S-R contingencies, or whether they are a consequence of awareness and the hypotheses (Dulany, 1961) held by the subject as to what he should do in order to obtain reinforcement. While it has appeared that verbal conditioning is only found in those subjects who were inferred to be aware of the S-R reinforcement contingencies (Spielberger, 1965), this has not been true in all cases (Farber, 1963), and an element of ambiguity remains since (*a*) judgments of awareness have been shown to be dependent upon the thoroughness of the postexperimental interview (Levin, 1961), and (*b*) it is possible to argue that awareness may be a consequence of conditioning rather than a cause (Postman & Sassenrath, 1961). The present experiment attempts to reduce these ambiguities through experimental manipulation of awareness: it investigates the extent to which learning curves are a function of awareness.

---

[1] This study was supported in part by Grant MH 11081 from the National Institute of Mental Health, United States Public Health Service. The author wishes to express his thanks to Martha Henderson, Estelle Resnick, and Steve Barkley for their excellent performance in their respective roles in this experiment.

Methodically, the use of verbal conditioning provides a convenient vehicle for the study of the subject in his role as expert witness. Orne (1962) has presented persuasive evidence for the effects of the "demand characteristics" of experiments on the subject's performance. Simply put, these are the effects of the subject's hypotheses or beliefs about the nature of the experiment and his attempt to be a "good subject" and provide the data sought for by the experimenter. But because not all subjects necessarily perform as they think the experimenter wishes them to (Farber, 1963), it seems reasonable to refer to those who do, or those who have been subjected to some experimental treatment which may be expected to increase the likelihood that they would, as "beneficent subjects." The verbal conditioning task permitted the author to manipulate awareness directly by either informing or not informing subjects how the experimenter wanted them to perform. Because of certain sociocultural aspects of the experimental situation, to be described below, it seemed reasonable to assume that the informed subjects would also be beneficent. Thus it was possible to compare not only their performance in verbal conditioning as a function of awareness and beneficence, but also the quality of their testimony in the postexperimental awareness interview. If demand characteristics enter into the subjects' performance in the experiment proper, there seems to be no reason why they would not also affect their performance as expert witnesses in the postexperimental interview.

In summary, the purpose of this experiment was to contribute data relevant to the question of the role of awareness in verbal conditioning and also to study the behavior of subjects in postexperimental awareness interviews, where both awareness and demand characteristics have been experimentally manipulated.

METHOD

Subjects. The subjects were 32 male students enrolled in an introductory psychology course for which participation in four experiments was a requirement.

Stimulus materials. A Taffel (1955) conditioning procedure was used for which 100 3 × 5-inch white index cards, each containing a past-tense verb and the pronouns I, WE, YOU, HE, SHE, and THEY, provided the stimuli. The verb was typed in the center of the card and the pronouns beneath it, in a different sequence from one card to the next.

Procedure. The subjects were randomly assigned in equal numbers according to a predetermined schedule to either an informed or an uninformed group. The experimental personnel consisted of a male undergraduate confederate, whose job it was to either inform or not inform each subject of what the experiment was all about; a female graduate student, who acted as the experimenter and whose physical endowments and manner were such as to leave little doubt concerning her ability to elicit beneficence from the typical male undergraduate; and an undergraduate female who conducted the postconditioning interview.

Each subject was identified throughout the experiment by a code number based upon his order of arrival for the experiment. The schedule followed by the confederate which determined whether or not a subject was to be informed was not known to either the experimenter or to the interviewer. Thus each subject was run and interviewed blind so far as knowledge of which group he was in. Additionally, the interviewer had no knowledge of the subject's performance in verbal conditioning at the time that she conducted her awareness interview.

Upon arriving for the experiment each subject found the confederate sitting outside the laboratory and a vacant chair beside him. A sign by the laboratory door requested all subjects for the experiment to be seated until called by the experimenter. The confederate engaged in small talk with all subjects so that when the experimenter arrived to bring the subject into the lab she found them all engaged in conversation and could not determine which group they were in. For those subjects in the informed group, the confederate began his conversation with the following remarks delivered in a well-rehearsed, *entre nous* fashion:

> Are you waiting to be in the experiment "words?" I just got through with it and I have to wait for a second part or something. I think the girl who is running it is doing it for her doctor's dissertation. Boy I'll bet she's worried about getting the right results. It's a funny experiment. You have to sit there and make up sentences using these words she has on a card and it seems she wants you to make up sentences using *I* or *We* as the pronoun. I think I did pretty well once I caught on. Better not say that I told you about it.

In this way awareness was provided the informed group, and, given the aforementioned characteristics of the experimenter, it seemed likely that these subjects would also be beneficent.

Subjects were run individually by the experimenter, who gave them the conventional Taffel instructions to the effect that they were to make up a sentence for each card using the verb and one of the six pronouns beneath it, saying the first sentence that came to mind. The experimenter, who was separated from the subject by a small wooden partition resting on a table, presented the stimuli on a stimulus holder in front of the subject, recorded his pronoun choice, and, during the conditioning trials, reinforced each sentence beginning with *I* or *We* by saying "good." The first 20 sentences, which the subjects were told would be for practice, were used to determine the subject's operant level, and the remaining 80 constituted the conditioning trials. The experimenter knew that half of the subjects were in each group, and immediately after the subject left the room recorded her guess as to whether or not he had been informed.

Following completion of conditioning the experimenter thanked the subject for his cooperation, asked him not to tell anyone else about the experiment, and then led him into another room in the laboratory suite and introduced him to the interviewer who, she said, wished to ask him a few questions. The interviewer then administered a 12-question postconditioning interview for awareness used by Spielberger and Levin (1962). The questions in this interview began in a nonleading way with, "Did you usually give the first sentence that came to your mind?," and became increasingly direct and to the point with the last question being, "While going through the cards, did you think that her saying 'good' had anything to do with the words you chose to begin your sentences? What?" When awareness was judged on the basis of the subjects' answers to the entire interview, Spielberger and Levin found a highly significant difference in the number of conditioned responses between aware and unaware subjects, and no difference between unaware subjects and those in a nonreinforced control group. Our interest in using this interview was not only as a means of inferring awareness; by noting the question number on which the subject indicated awareness it seemed also to offer a way of investigating the effects of demand characteristics and beneficence upon the subject's candor.

If the subject reported awareness at any point during these 12 questions he was then asked to identify the trial block in which he first became aware. Finally, as the last question all subjects were asked: "Did you know anything about the experiment before you participated in it?" The interviewer recorded all answers verbatim and the early interviews were tape-recorded as a check on the interviewer's performance. The interviewer proved so competent that it was not felt necessary to tape all of her interviews.

RESULTS

Interview data and detection of beneficence. The experimenter's ability to identify which subjects had been informed and which had not was evaluated by comparing her judgments against their actual treatment. She correctly identified 56% of the informed subjects and 75% of the uninformed subjects, but her overall hit rate of 66% was not any better than expected by chance ($X^2 = 2.07$, $df = 1$). Hence, the experimenter was not able to detect the beneficent subject on the basis of his performance alone.

A subject was classified as aware if any of his answers during the postconditioning interview could be taken as statements of the reinforcement contingency of "good" for sentences beginning with *I* alone, with *We* alone, or with *I* or *We*. Subjects were classified without knowledge of their preexperimental treatment or their performance in verbal conditioning. By definition it would be expected that a larger proportion of the informed subjects would be aware than would the uninformed. Actually, 81% of the informed subjects were classified as aware, compared to 44% of the uninformed subjects. Thus the subjects' verbal reports did reflect their preexperimental treatment with some degree of accuracy ($X^2 = 3.33$, $df = 1$, $p < .05$, one-tailed). There were no differences between the aware subjects in the two groups, however, in the trial block which they designated as the one on which they first became aware of the contingency (informed, $M = 2.2$; uninformed, $M = 2.1$).

The number of questions it was necessary to ask in the postconditioning interview before the subject gave an answer from which awareness might be inferred was taken as a measure of his readiness to report awareness or his candor. While it might seem that it should take relatively little persistence to elicit a report of awareness from an informed subject since he ostensibly came into the experiment with it, considering the fact that he possesses guilty knowledge—the confederate asked him not to tell the experimenter that he had told him about the experiment—it would seem reasonable to predict on the basis of his possible conflict over this knowledge, as well as his beneficence toward the experimenter, that he would be more reluctant to reveal awareness as compared with an innocent uninformed subject. Giving each aware subject a score based upon the number of questions asked before he revealed awareness, the mean for the 13 aware, informed subjects was 5.3 and for the 7 aware, uninformed subjects, 3.3. This difference was evaluated by a procedure suggested by Edwards (1960, pp. 106–107) for unequal *n*'s and found to be statistically significant ($t = 2.02$, $p < .05$, one-tailed). Thus, as compared with the innocent subject, the beneficent subject appears less likely to give evidence revealing awareness unless pressed for it.

It will be recalled that the last question asked of all subjects was whether they had any knowledge of the experiment before taking part in it. Only 1 of the 16 informed subjects reported that he had been fully informed by the confederate, and 3 others only that the confederate had told them that they would have to make up sentences but nothing more. It is noteworthy that one of the partial reporters also said that he thought the interviewer might have known about the confederate—apparently he didn't want to be considered dishonest. Thus, 75% of the informed group denied any prior knowledge of the experiment, and *none* had spontaneously volunteered his knowledge to the experimenter during the experiment.

Verbal conditioning. The 100 trials were divided into five blocks of 20 each. Because the groups were not found to differ significantly in their frequency of *I-We* responses during the operant period, all analyses included this period as the first block of trials and were based upon frequencies rather than difference scores. These data are presented in Figure 1 for all subjects in the informed and uninformed groups and separately for only the aware subjects in these groups.

An analysis of variance comparing frequency of *I-We* responses for all subjects in the informed and uninformed groups indicated that the informed group was superior to the uninformed group ($F = 6.06$, $df = 1/30$, $p < .05$). It also produced a significant main effect for trials ($F = 13.19$, $df = 4/120$, $p < .01$), but no significant interaction between groups and trials, confirming the impression gained from Figure 1 that both groups showed the same rate of increase in frequency and differed only in level. A similar analysis conducted only on the performance of the aware subjects in each group showed the informed subjects still tending to be superior to the uninformed subjects ($F = 3.56$, $df = 1/18$, $p < .10$), with a highly significant effect for trials ($F = 17.49$, $df = 4/72$, $p < .01$), but again no significant Groups × Trials interaction. Thus, informed subjects inferred to be aware of the response-reinforcement contingency appear to differ from their uninformed counterparts only in their level of performance and not in the shape of their acquisition curve.

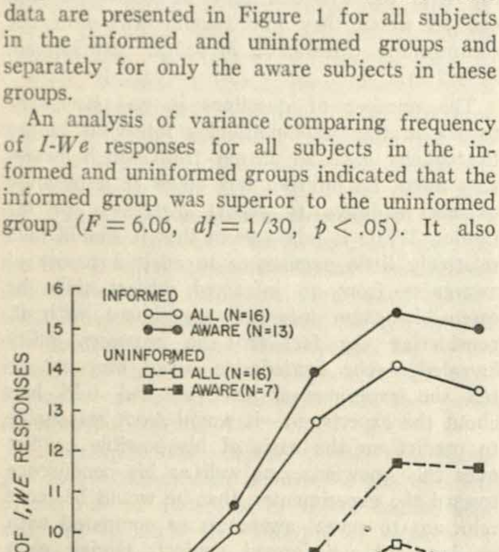


FIG. 1. Mean frequencies of *I-we* sentences in blocks of 20 each.

DISCUSSION

Because the rationale in so many human experiments assumes a naïve subject to begin with, perhaps the most common operational thread running through human research is the request made by the experimenter just before the subject leaves the experiment that he not tell anyone about it. How often this request is honored has seen little investigation, but it seems unlikely, college students being what they are, that it is anywhere near 100%. For this reason many experimenters also ask their subjects whether they knew anything about the experiment before taking part in it. It seems to be common practice to accept the subject's word on this. The data cast serious doubt on the validity of this practice.

For a variety of social psychological and other reasons, well discussed by Orne (1962) and others, the finding that 75% of the subjects in the present experiment did not give any indication that they had been told anything about it is not too surprising. But it is distressing. Given a beneficent subject, neither naïveté nor complete candor may be assumed.

These data strongly suggest that when subjects' verbal reports provide the evidential basis for hypothesis testing, demand characteristics may be expected to enter no less than when their performance, verbal or otherwise, is the source of data. For this reason studies in verbal conditioning, particularly of the Taffel type, must always be doubly suspect with respect to demand characteristics when the question is one of learning without awareness. Although more subjects in the informed than in the uninformed group indicated awareness during the postconditioning interview, it required a larger number of questions on the average to elicit this indication from them. Indeed, had they been subjected to one of the briefer types of awareness interviews, not at all uncommon in verbal conditioning studies (Levin, 1961), less than half of the informed aware subjects would have been so classified.

By definition, awareness implies the ability to verbalize. But although awareness may be a necessary condition for verbalization, it is obviously not a sufficient one. Nevertheless, the present study should not be taken to mean either that no faith may be put in subjects' verbalizations or that awareness is beyond the reach of scientific investigation. (It was true that those subjects who were classified as unaware, in the informed as well as the uninformed group, showed very little change in performance as compared with aware subjects.) Rather, it demonstrates that these verbalizations, like any other form of behavior, cannot be adequately understood without knowledge of their antecedent conditions. Like any other hypothetical construct, the meaning of awareness resides in the nomological net of which it is a part (Cronbach & Meehl, 1955) and in the corresponding converging operations (Campbell & Fiske, 1959; Garner, Hake, & Eriksen, 1956) by which it is defined. But there is no intrinsic reason, as Dulany (1962), as well as the present experiment have shown, why the study of awareness need place the experimenter entirely at the mercy of his subjects' testimony, whether they be beneficent, innocent, expert, or otherwise. Converging operations are possible and the study cannot be based upon subject's verbal reports alone. For the present experiment makes it obvious that to rely upon the subject as an expert witness would be to betray as much naïveté on the part of the experimenter as that which he hopes exists in his subject; without corroborative, independent data the subject's verbal testimony possesses no greater (or lesser) validity than does an unstandardized test.

Given that a subject is aware of the reinforcement contingency or what the experimenter wishes him to do, as Dulany (1962) and Farber (1963) have observed, it does not necessarily follow that he will behave accordingly. The verbal conditioning data appear to be consistent with these observations. Although all of the informed subjects in the present experiment had been provided with information which would have permitted them to obtain 100% reinforcement from the first trial onward, none did. While the first block of unreinforced trials might have been confusing and account for part of this, the gain in *I-We* responses in the informed group during conditioning was of the same order as that for the uninformed group, only their level of responding was higher. Awareness, whether defined by directly informing subjects or by inference from interview protocols, appeared to be associated with the same kind of growth curve commonly taken as reflecting the continuous incremental effects or reinforcement upon habit strength. But to interpret these curves in this fashion seems obviously inappropriate. For this reason, the conditioning data of this experiment would seem to have important implications for the conceptualization of the processes involved in verbal conditioning.

Most generally, they tend to make suspect interpretations of verbal conditioning data within an operant conditioning framework. While it would certainly be gratuitous to argue that the curves obtained in the present study simply reflect the subjects' beneficent responses to the demand characteristics of the experiment, it would be just as unfounded to view them as representing the operation of the same principles of learning and conditioning as believed operative in rats and pigeons. And if this is true in the present experiment, it may well be true in others. Thus it seems obvious that the shape of a response curve plotted over trials is, by itself, a weak basis upon which to infer the nature of the underlying process.

The great similarity between the curves for the informed and uninformed groups, while startling, is not uninterpretable. That the preexperimental treatment of informing or not informing subjects had an effect upon them is obvious from both the difference in height between the two curves, and the differences found in the number of aware subjects in each group and their readiness to admit awareness. However, when Dulany (1962) instructed subjects in the correct response and the reinforcement contingency their response rate increased abruptly; in the present experiment with essentially the same kind of information it did not. One possible explanation for this may be that in Dulany's experiment such an abrupt increase was socially sanctioned; the subject was acting upon legitimately obtained information. In the present experiment it is possible that the informed subjects were deterred from acting upon their knowledge and showing a similar abrupt increase in response frequency because of their fear that it would reveal their illicitly obtained information. Thus, one interpretation of the shape of the curve for the informed group is that it reflects the rate of change among the subjects' in their decisions to act consistently with their knowledge in order to obtain reinforcement or to satisfy the experimenter. And if this is true for the informed group, it is also a plausible explanation for the similarly shaped curve produced by the uninformed group. Although their awareness was not based upon ill-gotten knowledge, when they obtained it during the conditioning series it may be expected to have varied from one subject to another just as when they decided to act upon it.

What this suggests, of course, is that the curves obtained in verbal conditioning studies may reflect the operation of cognitive and volitional processes, or hypotheses and intentions as Dulany (1962) would have it, rather than the direct, automatic, and unconscious consequences of reinforcement. If this is the case, then it would seem to be misleading to speak of these studies as demonstrating verbal conditioning. They demonstrate the modification of verbal behavior to be sure, but many of them may owe their data more to the beneficence of their subjects than to the potency of their theory or reinforcement. This is not to say that reinforcement is unimportant. It provides the subject with information; it tells him what to do, albeit in a most inefficient way, and he may do it if he wishes to. But if he does, the question remains whether he should be considered conditioned or beneficent.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychological Bulletin,* 1957, **54,** 383–405.

BUCK, R. Comments on Buchwald's "verbal utterances as data." In H. Feigl & G. Maxwell (Eds.), *Current issues in the philosophy of science.* New York: Holt, Rinehart & Winston, 1961. Pp. 468–472.

CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 1959, **56,** 81–105.

CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychological Bulletin,* 1955, **52,** 281–302.

DULANY, D. E. Hypotheses and habits in verbal "operant conditioning." *Journal of Abnormal and Social Psychology,* 1961, **63,** 251–263.

DULANY, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. Eriksen (Ed.), *Behavior and awareness—a symposium of research and interpretation.* Durham, N. C.: Duke University Press, 1962. Pp. 102–129.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

FARBER, I. E. The things people say to themselves. *American Psychologist,* 1963, **18,** 185–197.

GARNER, W. R., HAKE, H. W., & ERIKSEN, C. W. Operationism and the concept of perception. *Psychological Review,* 1956, **63,** 149–159.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychological Bulletin,* 1958, **55,** 148–171.

KRASNER, L. The therapist as a social reinforcement machine. In H. H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy.* Vol. II. Washington, D. C.: American Psychological Association, 1962. Pp. 61–94.

LEVIN, S. M. The effects of awareness on verbal conditioning. *Journal of Experimental Psychology,* 1961, **61,** 67–75.

ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17,** 776–783.

POSTMAN, L., & SASSENRATH, J. M. The automatic action of verbal rewards and punishments. *Journal of General Psychology,* 1961, **65,** 109–136.

SPIELBERGER, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness—a symposium of research and interpretation.* Durham, N. C.: Duke University Press, 1962. Pp. 71–101.

SPIELBERGER, C. D. Theoretical and epistemological issues in verbal conditioning. In S. Rosenberg (Ed.), *Directions in psycholinguistics.* New York: Macmillan, 1965. Pp. 149–200.

SPIELBERGER, C. D., & LEVIN, S. M. What is learned in verbal conditioning? *Journal of Verbal Learning and Verbal Behavior,* 1962, **1,** 125–132.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology,* 1955, **51,** 496–501.

THORNDIKE, E. L. *The fundamentals of learning.* New York: Teachers College, 1932.

VERPLANCK, W. S. Unaware of where's awareness: Some verbal operants—notates, monents, and notants. In C. W. Eriksen (Ed.), *Behavior and awareness—a symposium of research and interpretation.* Durham, N. C.: Duke University Press, 1962. Pp. 130–158.

WILLIAMS, J. H. Conditioning of verbalization: A review. *Psychological Bulletin,* 1964, **62,** 383–393.

(Received July 15, 1966)

Volume 6 August 1967 Number 4

Diary No. 95
10-9-70
File No Lib
Bureau Edn. Psy. Research

Journal of Personality and Social Psychology

Editor: Daniel Katz, *University of Michigan*

Associate Editor: Robert B. Zajonc, *University of Michigan*

Robert P. Abelson
Yale University
Joseph B. Adelson
University of Michigan
Gordon W. Allport
Harvard University
Norman H. Anderson
University of California, San Diego
Elliot Aronson
University of Texas
Albert Bandura
Stanford University
Leonard Berkowitz
Cornell University
James Bieri
University of Texas
Jack W. Brehm
Duke University
Donald R. Brown
University of Michigan
Eugene Burnstein
University of Michigan
Arnold H. Buss
Rutgers University
Launor F. Carter
Systems Development Corporation
Rosalind D. Cartwright
University of Illinois
Medical Center, Chicago
Dwight W. Chapman, Jr.
Vassar College
Irvin L. Child
Yale University
Nickolas B. Cottrell
Kent State University
William Dember
University of Cincinnati
Morton Deutsch
Columbia University
Seymour Feshbach
University of Colorado
Donald W. Fiske
University of Chicago
Basil S. Georgopoulos
University of Michigan
Edwin P. Hollander
State University of New York at Buffalo
Murray Horwitz
New York University
Herbert C. Kelman
University of Michigan
Daniel J. Levinson
Connecticut Mental Health Center, New Haven
Richard A. Littman
University of Oregon
Robert B. MacLeod
Cornell University
Melvin Manis
Veterans Administration Hospital
Ann Arbor, Michigan
David C. McClelland
Harvard University
Charles G. McClintock
University of California, Santa Barbara
William J. McGuire
Columbia University
Vincent Nowlis
University of Rochester
Harold L. Raush
University of Michigan
Milton E. Rosenbaum
University of Iowa
William A. Scott
University of Colorado
Janet Taylor Spence
University of Texas
Percy H. Tannenbaum
University of Wisconsin
Harry C. Triandis
University of Illinois
Joseph Veroff
University of Michigan

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

Part 1 of 2 Parts

The *Journal of Personality and Social Psychology* is devoted to basic research and theory in the broad field of social interaction and its determinants and effects. More specifically, it is concerned with the subareas of (*a*) social interaction and group process, (*b*) interpersonal perception and attitude change, (*c*) the psychological aspects of formal social systems and less structured collective phenomena (e.g., the public opinion process and mass movements), (*d*) the psychological analysis of cultural forms and products (e.g., language), (*e*) the socialization process both at the child and adult levels, (*f*) social motivation and personality dynamics, (*g*) the structure of personality, and (*h*) the relation of personality to group process and social systems. Animal studies as they relate to an understanding of social processes are definitely appropriate for the journal. Methodological articles and studies of the development of personality instruments are not excluded but are considered marginal unless presented as part of a substantive piece. Where existing journals already serve a given specialized interest within the broad domain of the field, articles appropriate to this area of specialization will have low priority for the *Journal of Personality and Social Psychology*.

The JOURNAL is issued monthly, three volumes per year. The subscription price is $30.00 annually (foreign, $30.50). Single issues are $3.00.

Manuscripts must be submitted in duplicate and should be sent to the Editor-Elect:

Dr. William McGuire
Journal of Personality and Social Psychology
Department of Psychology
University of California, San Diego
P. O. Box 109
La Jolla, California 92037

Manuscripts must be accompanied by an abstract of 100–120 words. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or may be obtained from the Editor or from the APA Central Office.

Authors submitting manuscripts are requested to use full journal titles in reference lists; do not abbreviate.

The following policies govern reprinting of materials copyrighted in APA journals: (*a*) approval is required for reprinting tables and figures, and for text only if more than 500 words (total from one article) in length; (*b*) approval is required for reprinting articles in whole or in major part, and is contingent on the author's approval; (*c*) negotiation, where possible, for the dedication of royalties from commercial publishers to the American Psychological Foundation.

HELEN ORR
Managing Editor

JOANNE BERGEN
Technical Editor

HENRY S. SALMON
Advertising and
Business Manager

ROBERT D. GORMAN
Subscription Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

Changes of address must reach the Subscription Office by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Sts., Lancaster, Pa. 17604
and 1200 Seventeenth St., N. W., Washington, D. C. 20036

Second-class postage paid at Lancaster, Pa.
Copyright © 1967 by the American Psychological Association, Inc.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 371-380

EFFECTIVENESS OF DEBRIEFING FOLLOWING DECEPTION EXPERIMENTS

ELAINE WALSTER
University of Rochester

ELLEN BERSCHEID, DARCY ABRAHAMS, AND VERA ARONSON[1]
University of Minnesota

This study was designed to test the notion that deception experiments might be especially disturbing to a person already concerned about the area involved in the deception. Specifically, it was hypothesized that if a person who was extremely concerned about his adequacy in social situations was given false information about his adequacy as part of an experiment, it might not be easy to debrief him at the end of the experiment. In this experiment, degree of concern about one's social skills was varied in 2 ways: concern was manipulated; high-concern and low-concern Ss were selected on the basis of a personality test. Regardless of the way in which concern was varied, there is no evidence that high-concern Ss are more difficult to debrief than low-concern Ss. Unfortunately, however, there is evidence that (a) regardless of degree of concern, debriefing may not be effective immediately, and (b) debriefing may be ineffective, even after a longer time delay, for certain personality types.

In order to investigate many social psychological questions, it is often necessary to deceive subjects within the experimental context. Deception generally serves two purposes: First, in many experiments the experimenter's hypothesis would be all too apparent to the subjects if a false explanation of the experimenter's purposes were not provided. If the purpose of the experiment were clear to subjects, they might try to assist or thwart what they believe to be the experimenter's aims (Orne, 1965). A false explanation eliminates this problem, for even if subjects try to help or hurt the experimenter, they will in fact be responding to an irrelevant hypothesis. Second, deception might be required in order to manipulate the independent variable. Such deception permits the laboratory study of variables which would otherwise not be accessible to investigation.

In spite of, or perhaps because of, the regularity with which deception is used in experimentation, several reservations have been expressed about the use of this experimental technique (e.g., Kelman, 1965; Orne, 1965; Silverman, 1965). Critics of deception experiments seem to be primarily concerned with the possibility that deception might cause the subject permanent harm. In partial response to this criticism, defenders of the deception technique have pointed out that subjects in deception experiments are almost always thoroughly debriefed at the end of the experiment. In the debriefing, the subjects are usually told the nature of the deception, the true purpose of the experiment, and the reasons why deception was necessary. Kelman (1965) suggested that "in general, the principle that a subject ought not to leave the laboratory with greater anxiety or lower self-esteem than he came with is a good one to follow." Most experimenters seem to operate under the conviction that a thorough debriefing will accomplish the goal of returning the subject to his preexperimental state.

Yet the fact that experimenters who employ deception usually take a great deal of time and care to debrief their subjects has not entirely allayed fears concerning the use of deception. The obvious question which remains is how can the researcher be certain that his debriefing has been successful? How can he be sure that when the subject walks out the door, he does not do so with lower

---

[1] The authors would like to express great appreciation to Bill Walster and Raymond O. Collier (who analyzed our data), and to Susan Boye, Shirley Damrosch, David Landy, and Darwyn Linder, who helped to conduct the studies reported in this paper and wish to acknowledge the many helpful comments and suggestions made by Dr. John M. Darly. This study was supported in part by National Science Foundation Grant GS 1056.

self-esteem than when he came in? In spite of the fact that it is of crucial importance to experimenters that debriefing remove the effects of the deception, there is no experimental evidence that debriefing, in fact, does so.

Deception experiments differ so greatly from one another in the nature and degree of deception used that even the harshest critic of this technique would be hard pressed to state unequivocally that all deception has potentially harmful effects. There are, however, two frequently mentioned dangers of deception experiments, to which some experiments are more liable than others. First, some critics have voiced their concern that lying to people may lead them to lose faith in their fellow human beings. Because scientists are ordinarily highly respected, the discovery that a scientist will lie might upset subjects even more than lies told by others. Secondly, and perhaps more importantly, it has been pointed out that some deception manipulations are emotionally disturbing to a subject, and that some disturbances might not be entirely amendable by debriefing. Consider, for example, an experiment dealing with the differences between task performance of high self-esteem and low self-esteem people. High self-esteem might be manipulated by telling subjects that they are virile, creative, and personable individuals. Low self-esteem might be manipulated by telling subjects they have negative qualities. It is evident that for the duration of the experiment the subjects in the low self-esteem condition are likely to be upset. But what about *after* they are debriefed? It has been suggested by Kelman that if one is a normal person, it is probable that such a shock to one's self-concept can be overcome and the assurances of the experimenter in the debriefing session can be readily accepted. However, Kelman points out that there are undoubtedly instances in which simply telling the subject that the results of his virility-creativity-personableness test were falsified will not erase the effects of the deception.

Under what conditions might we expect debriefing to fail? Suppose that a subject is told that his test results indicate that he is not very creative. In fact, the experimenter reports, few people tested have ever scored so low on creativity. Further suppose that this subject happened to be a budding poet who picked up his mail on his way to the experimental session and discovered the fourteenth publisher's rejection slip for his first serious effort. While pursuing his way to the experiment, the subject might quite naturally wonder if the series of rejections should be attributed to his lack of talent or to the possibility that the uncultured masses are not clamoring for sonnets about the Crimean War.

It seems quite possible that the experimenter's authoritative evaluation of the subject's creative talents would initiate in this particular subject some independent thinking during the course of the experiment. It is quite likely, for example, that this subject would try to reach some consistency between the content of the experimenter's report and his own original ideas about his level of creativity. To do this, he might well reexamine his self-concept and selectively recall past incidents, most notably the 14 rejection slips, which would agree with the experimenter's information. Memories of criticisms from friends and family, the recollection of some low grades in English composition, would, when interpreted in the light of the test results, strengthen his belief in his supposed low creativity. Consequently, at some point in the course of the experiment, the subject might decide that since his own cognitions augment the experimental evidence, what the experimenter said was true. He might even come to the conclusion that he himself had been imperceptive not to realize his lack of creativity before.

At the completion of the experiment, of course, the subject-poet would be informed that the negative evaluation he received was chosen at random, and it would be explained that it was just as likely that he could have received a neutral or favorable evaluation of his creative talents. As previously mentioned, there is ordinarily little reason for the subject not to accept completely the notion that he has been deceived, that he is not the unimaginative dullard he thought he was, and be none the worse for wear. It even seems quite probable that the subject might believe the debriefing message in its entirety, that is, that the creativity test was not genuine. In this case, however, his own supporting and freshly organized cognitions might remain. It is still true that he *has* gotten 14 consecutive rejection slips and that he *did* receive those low grades. Consequently, it is possible that, though the specific anxieties produced by the deception might be completely removed by the debriefing, his general opinion about himself might well be lowered, and his life turned upon a new course, because of the extra thinking the manipulation initiated.

When the deception happens to strike an area of deep concern and worry to the individual, when it is likely to initiate a train of thought which would not be altered by the revelation of the deception, it is possible that the damage done to a subject by the deception might be irreversible.

The authors will report in this paper an experiment which tests the hypothesis that it will be more difficult to successfully debrief (i.e., return to his preexperimental state) a subject who has received false information on some aspect of himself about which he is currently concerned, than it will be to debrief a subject who has received information which is irrelevant to his current concerns. While the preceding discussion focused on the possible residual effects of receiving negative information, a parallel result could be expected to occur when the subject has received positive information. That is, it might also be more difficult to debrief someone who receives positive information in an area of current concern. This experiment, then, was designed to test for the existence of both positive and negative residual effects after debriefing.

METHOD

Overview

Three steps were required to test our hypothesis. It was necessary that: (*a*) some subjects be very concerned about their characteristics in a given area while other subjects be unconcerned. Subjects either could be selected because they did or did not manifest this concern, or concern could be experimentally manipulated. We chose both to measure preexperimental concern and to manipulate concern experimentally; (*b*) subjects participate in a deception experiment in which false information relevant to the area of concern is provided; (*c*) subjects have time to think about the false information and an opportunity to use it to organize events in their past life. We, of course, expect subjects to be more inclined to relate the false material to previous life experiences when they are deeply concerned about the area than when they are not; (*d*) subjects be thoroughly debriefed and, following the debriefing, success or failure be measured.

Procedure

The 80 subjects who participated in this study were freshmen and sophomore women enrolled at the University of Minnesota.[2] Subjects were recruited from an introductory psychology course and from the university library. All subjects had taken the MMPI as part of the freshman testing program.

In an initial contact, subjects agreed to participate in two separate experiments. They were told that the first experiment would have an hour delay between the first and second part. Thus, a second experiment had been scheduled during this hour for their convenience. They were told they could either participate in this experiment or not as they chose. In fact, all subjects chose to participate in both experiments. In reality, of course, the two experiments were both parts of the same experiment.

The purpose of Experiment 1 was twofold: (*a*) to *measure* the subject's preexperimental concern about the kind of social impression she makes, and (*b*) to randomly assign the subject to an experimental group and *manipulate* her concern about the kind of social impression she makes. We wanted half of our subjects to be highly concerned and curious about their social abilities and the other half to be little concerned about these abilities.

Subjects arrived at Experiment 1 in groups of four, and each subject's concern about her social abilities was measured as soon as all had arrived at the experimental room. Each subject completed the Fear of Rejection Scale. According to Rosenfeld (1964), this scale measures one's concern or fears about social rejection.

The remainder of Experiment 1 was designed to manipulate subjects' concern about their social skills. This was done in the following way: Experimenter 1 told the subjects that the purpose of Experiment 1 was to find out "something about how people make first impressions and how they react to a first-impression situation." She explained to them that two college seniors would interview each subject for about 10 minutes, asking her several questions. Following the interview, the seniors would evaluate the subject. The seniors were described as outstanding students, both academically and socially, who had been doing such interviews for about a year. The subjects were then introduced to our two interviewers, a very attractive male and female.

---

[2] Sixteen of these subjects were discarded from our analysis: 1 subject had recently received an MMPI report in counseling, 1 subject could not understand the procedure, 2 subjects failed to answer all questions, 1 subject was very suspicious that we were deceiving him, and 11 subjects were slightly suspicious of the procedure for reasons they could not specify. We discarded any subjects who were in the least suspicious. We did, however, reanalyze the data including all subjects, and it is clear that including the data of the suspicious subjects would not alter our conclusions in any way.

Subjects were interviewed one at a time, while the other subjects continued filling out forms in an adjoining room. The interviewers stood behind a one-way mirror, so that the subject could not see their facial reactions. It was the confederates' job, by the tone of their voices and their comments, to make each subject very unsure about how well she had done. In pretests the authors selected several interview questions which were difficult for subjects to answer (e.g., "How do you feel towards people who are better looking than yourself?" or "What is your favorite joke?"). The two confederates by their manner approved some answers and disapproved of other answers.

At the completion of each subject's interview, the experimenter checked to see whether the subject had been randomly assigned to a group that should have *high concern* with how they had done in the interview situation or to a group that should have *low concern* with their performance. If the subject had been assigned to a *high-concern* group, the experimenter made several comments designed to impress the subject with the fact that her performance in the ambiguous interview situation was extremely important. To point out how important interviews are in daily life, the experimenter cited statistics concerning the importance of first impressions. She explained that since automation could take over many routine jobs but could replace only a few jobs requiring personal contacts, the ability to create a good first impression would be even more important in the future. Finally, the experimenter attempted to refute any arguments that the subject might come up with to challenge her claim that the first impression she made in the interview situation was important. She noted that some subjects had felt that because they had not been warned in advance that the interview was important, or because the interview was too short, or because they were nervous, the impression they made was not a valid one. Experimenter 1 explained that since these factors were present in real life interviews, they actually increased the validity of the evaluation. If the subject had been assigned to the low-concern condition, Experimenter 1 made totally different comments after her interview. She explained that although her performance would be of interest, one could not tell too much about a person on the basis of just one interview. She said such a situation was not a very good predictor of how one did in real life and commented that at present she was just pretesting the procedure. Finally, she presented all the reasons why a person might argue that her interview performance was not a valid indicator of her personality, but this time she agreed that these objections were valid ones.

When all four subjects had completed their interviews, Experimenter 1 explained that during the next 20 minutes the interviewers would be busy evaluating the subject's performance. She told the subjects that they could leave for the second experiment, but that those who had given an unusually good or an unusually poor performance would be called back to Experiment 1 for more extensive interviewing after Experiment 2 had been completed. The remaining subjects would come back simply to complete a brief questionnaire. She said that she would send a note to them sometime during the second experiment, telling them whether to plan on a long interview or a short one.

The purpose of Experiment 2 was to lead one-half of the unconcerned and one-half of the concerned subjects to believe for the duration of Experiment 2 that they possessed good social skills, and the remainder of the subjects to believe that they possessed poor social skills. To impart these beliefs to subjects, Experimenter 2 told them, when they arrived at Experiment 2, that she was interested in several aspects of sociability or "the ability of a person to enter into warm, long-lasting and meaningful relationships with other people." She said she already had secured MMPI sociability scores for all of them. After briefly discussing the omniscience of the MMPI's sociability scale, the experimenter told the subjects that many researchers were now engaged in developing a shorter test than the MMPI to measure sociability, and also a test more easily utilized with youthful test takers. The experimenter explained that today they would take a newly developed test, the "Social Aptitude-Achievement Test." The test booklet which the experimenter then proceeded to pass out was designed simply to appear to be a valid measure of sociability. It presented several social situations in which one might find oneself and asked the subject how she would respond to each situation. After approximately ½ hour the experimenter collected the tests, and announced she would score them immediately so she could see to what extent the results agreed with those of the MMPI. Before leaving the room to score the test, Experimenter 2 gave each subject what was purported to be her MMPI report to read, while she was out scoring the tests "in case she was interested." Of course, all subjects were interested.

The MMPI report which the subject received was false. Only two alternate forms existed. By random assignment, one-half of the subjects were to receive a positive social assessment in their report. Their reports specified that the subject had scored near the ninetieth percentile on the "Sociability Scale" of the MMPI and contained a two-page personality assessment, supposedly written by a clinician at the University of Minnesota. This assessment indicated that they were "natural and balanced" in their relations with others, that they played the ascendant role in relationships with others without being overbearing, but that possibly they were not aware of the "potential richness of their skills" in the social area. Subjects assigned to the negative condition received profiles indicating that they had scored near the eighteenth percentile in sociability. The written portion of their profiles claimed that they were reluctant to become involved with others, that they could not express their real feelings to others, and that they lacked self-insight to the point of being largely unaware of these shortcomings.

The subjects were then given time to read and to think about these clinical reviews. It was the authors' expectation that those subjects who were very concerned about how they had performed in the interview in Experiment 1 would use the MMPI personality information in an attempt to resolve the ambiguity they felt concerning their interview performance. It was thought that unconcerned subjects would be less interested in their performance (which they had been told was not a good indicator of their real personality), and thus less concerned about resolving any ambiguity they felt.

To remind subjects of Experiment 1, and to reinforce the ambiguity of evidence concerning their interview performance, Experimenter 1 walked in the room while Experimenter 2 was out of the room scoring the sociability tests, and handed a message to each subject. Presumably, this note was to tell subjects whether or not they had given an unusual performance in Experiment 1 (and thus had to return for a long interview at the end of Experiment 2) or had done about average. Actually, each subject received a note saying that her performance had been unusual. Whether it was unusually good or unusually poor was not specified. Experimenter 1 then supervised subjects until Experimenter 2 returned.

When Experimenter 2 returned she handed each subject her test scores on the Social Aptitude-Achievement Test. In all cases the scores agreed with those provided in the false MMPI report. A subject assigned to the positive condition was told that she had received a percentile score of 94. A subject in the negative condition was told she had received a score of only 20.

Once subjects had had a chance to consider the deceptive material we had provided, we were ready to debrief them. However, before we could do so, a bit more deception, designed to increase the plausibility of the situation, was necessary. We wanted Experiment 2 to seem complete in and of itself to the subject. We certainly did not want any subjects to connect Experiment 2 with Experiment 1. If we had simply explained to subjects at this point that we had been lying to them about their personality reports, many of them would have wondered what we had accomplished by lying to them. Consequently, the next step was designed to provide a rationale for Experiment 2 and to make it seem complete in and of itself by the time of debriefing.

Thus, Experimenter 2 asked the subjects to read and evaluate two case reports. In these case reports, a great deal of personality information was provided about two girls. Interspersed with the information was the fact that one of the girls was very sociable and outgoing, while the other was a very introverted person, lacking in social skills. Subjects were asked to comment upon their own reactions to the girls. Specifically, they were asked to indicate on a series of 20-point scales how much they liked the girls, to evaluate various personality characteristics of the girls, and to indicate how similar each of the girls seemed to themselves. After the subjects had finished rating the two girls, the stage was set for the debriefing. Experimenter 2 explained that the MMPI and Social Aptitude-Achievement Test scores she had given the subjects had been entirely false. She explained that she had never seen the subject's MMPI test, and that in fact there was no such thing as an MMPI sociability scale nor a Social Aptitude-Achievement Test. Subjects were shown both the high sociability and the low sociability MMPI reports, and it was pointed out that which report the subject received had been determined by chance. This explanation took a minimum of 20 minutes. Experimenter 2 then briefly explained to subjects why they had received a false personality report. She told the subjects that the experiment was actually concerned with how high or low self-esteem individuals felt about other people with the same traits. She explained that by raising the self-esteem of some subjects and lowering the self-esteem of others and then asking them to rate similar or dissimilar stimulus girls, she could answer this question in a clear-cut way.

Any questions asked by the subjects were answered and the whole procedure was discussed with subjects until Experimenter 2 felt confident that subjects understood that the false personality reports they had received were not their own.

Measuring of Postdebriefing Effects

The relative effectiveness or ineffectiveness of our debriefing procedures was measured in the following ways:

1. Immediately after the false debriefing, Experimenter 2 explained that a knowledge of what the subjects' personalities were *really* like would be valuable to her in analyzing the data. Thus, subjects were asked to fill out a questionnaire which they believed to be anonymous. They were asked to describe how plausible the personality information they had received had been, and, more importantly, to rate themselves on several personality traits. The traits on which subjects rated themselves were: their "ability to make deep, close, lasting, mature friendships," their "popularity," their "sensitivity to the feelings of others," and their "naturalness and spontaneity with others." Subjects utilized 4-point rating scales. The higher the score, the more "sociable" the subject indicated she was. The subject's scores on each trait were summed and the total score gave us a *Sociability Index*.

2. After being released from Experiment 2, all subjects reported back to Experiment 1. Since all had been told that they had done "unusually" in Experiment 1, they all believed that they would be required to answer a great number of questions for Experimenter 1. Undoubtedly, most subjects assumed that the other subjects were "average," and were returning to Experiment 1 simply to answer a few questions. All subjects were given a questionnaire to complete. The first question asked, "How well do you think you scored on your interview in Experiment 1?" Question 2 asked, "How well do you generally do in first impression situations?"

Possible answers to these questions ranged from 0 ("I guess I did extremely poorly") to 18 ("I guess I did extremely well"). Answers on these two questions were combined to form the *Interview Performance Index,* a measure of the extent to which a subject felt that she had done well in the initial interview.

The final question on this questionnaire was essentially a manipulation check. It asked, "Is this [the interview] a good example of what you are like in real life?" Subjects could answer either "Yes" or "No." It will be recalled that subjects in the high-concern condition had been told that their interview performance was important since it was representative of them, while low-concern subjects had been told that their performance was unimportant and unrepresentative of them.

Subjects were debriefed at great length after all of these measures were collected. Since subjects viewed this experiment as an effort to find out the extent to which deception might or might not be harmful to them, virtually all of them indicated that they were happy to have participated.

RESULTS AND DISCUSSION

All of the F tests in this paper are $2 \times 2 \times 2$ analyses involving selected concern, manipulated concern, and type of personality report provided to subjects.

Manipulation Checks

In Experiment 1, we tried to make half of our subjects concerned with the quality of their interview performance, and half unconcerned with their performance. This manipulation appears to have been effective. On their final questionnaire, subjects were asked to state whether or not the interview situation was a good example of what they were like in real life. More of the manipulated high-concern than low-concern subjects felt that the interview *was* representative ($F = 3.40$, $df = 1/56$, $p = .07$).[3]

During Experiment 2, some subjects were led to believe that they had received high scores on two tests of sociability, and others were led to believe that they had received low sociability scores. Did our subjects accept this information? The evidence indicates they did. It will be recalled that as part of Experiment 2, subjects were asked how similar they were to each of the girls described in the two case reports. Girls in the high-sociability condition indicated that they were much more similar to the high-sociability stimulus girl ($F = 31.04$, $df = 1/56$, $p < .001$), and that they were much less similar to the low-sociability stimulus girl ($F = 80.13$, $df = 1/56$, $p < .001$) than did girls in the low-sociability condition.

Results

It was hypothesized that it would be more difficult to debrief the high-concern subjects than low-concern subjects. Thus, we expected high-concern subjects who were told they possessed good skills in Experiment 2 to overestimate their performance, even after debriefing, to a greater extent than would those low-concern subjects who were also told they possessed good social skills. Similarly, we expected high-concern subjects who were told they possessed poor social skills to underestimate their social skills, even after debriefing, to a greater extent than low-concern subjects.

In this experiment, subjects were classified as high concern in two ways: (*a*) *Selected concern:* concerned and unconcerned subjects were selected on the basis of their scores on Rosenfeld's Fear of Rejection test. Those subjects who scored from 35–65 on this test were classified as high in their concern about social rejection. Those subjects who scored 0–34 were classified as low in their concern about social rejection; (*b*) *manipulated concern:* concern was manipulated in Experiment 1 as has been previously described.

How much social skill a subject believed he possessed *after debriefing* was measured by his self-ratings on the Sociability Index.

From Table 1 it is clear that regardless of whether we deal with manipulated concern or selected concern, and regardless of whether we consider self-estimates on the sociability index or on the interview performance index, high-concern subjects do *not* seem to be more difficult to debrief than low-concern subjects. Selected concern and type of deception do not interact in affecting subjects' estimates of their sociability ($F = 1.52$, $df = 1/56$) or of their interview performance ($F = 5.76$, $df = 1/56$) as we predicted they would. Though the last

---

[3] The reader might feel that a multiple-contingency test is a more appropriate method of analysis (Sutcliffe, 1957). Such an analysis gives us almost identical results to those reported above ($\chi^2 = 3.33$, $df = 1$, $.10 > p > .05$). Recent research would indicate that with a reasonable sample size there is no appreciable difference between a chi-square or an F test on discrete data (Cochran, 1950).

TABLE 1

SUBJECTS' ESTIMATES OF THEIR SOCIABILITY AND INTERVIEW PERFORMANCE AFTER DEBRIEFING IN VARIOUS CONDITIONS

Degree of concern	Deceptive personality information given	*N*	Sociability index[a]	Interview performance index[b]
Selected concern				
High concern	Told sociable	16	9.86	20.54
	Told unsociable	20	8.70	17.62
Low concern	Told sociable	16	10.19	20.64
	Told unsociable	12	9.83	23.20
Manipulated concern				
High concern	Told sociable	16	9.93	20.70
	Told unsociable	16	8.93	19.78
Low concern	Told sociable	16	10.12	20.49
	Told unsociable	16	9.31	19.64

[a] A high score indicates the subject reported high ability to make friends, popularity, sensitivity, and spontaneity.
[b] A high score indicates the subject estimated that she did well in the initial interview and that she generally makes a good first impression.

interaction *F* is significant, the mean estimates are clearly not in the predicted direction. We will discuss this interaction later in the paper. When we consider manipulated concern, the results are the same. After debriefing, high-concern subjects do not seem to be more difficult to debrief than low-concern subjects, whether we consider the Sociability Index ($F = .12$, $df = 1/56$) or the Interview Performance Index ($F = .00$, $df = 1/56$).[4]

The data obtained from this experiment leads us to believe that our hypothesis is incorrect. This belief is further strengthened by information that we have not yet presented to the reader. The study presented here is not the first time that this hypothesis was tested. In a very lengthy pilot study conducted by Abrahams (1967) the hypothesis of this paper was tested with only minor variations. (Ostensibly the Abrahams study was concerned with "problem solving ability" and "creativity" rather than with "interview performance" and "sociability.") The data from Abraham's experiment also failed to provide any evidence that highly concerned subjects are more difficult to debrief than unconcerned subjects.

Although we feel that our hypothesis is incorrect, this does not mean that we have concluded that debriefing is uniformly effective with all subjects. Our data indicate that subjects in the various conditions do exhibit lingering aftereffects of the deception, although they are not the effects we predicted.

Let us first consider the estimates subjects made of their own sociability immediately after having been debriefed. For all subjects, it appears that the personality report had an impact which lasted at least for a few moments beyond the occurrence of debriefing. Regardless of whether concern was manipulated or selected, and regardless of level of concern, subjects who were given favorable sociability reports rated themselves significantly higher on the sociability index than did those subjects who received an unfavorable personality report (main effect $F = 6.12$, $df = 1/56$, $p < .05$). Since the type of sociability report a girl received was determined by chance, we must assume these differences are due to the fact that the subjects were not entirely disabused of the information they received in Experiment 2.

It is possible, of course, that these aftereffects are due to the fact that our experimenters were simply poor debriefers. However, this possibility does not reduce our concern about these results for two reasons: (*a*) In the Abrahams (1967) experiment four different experimenters conducted the variation of the study reported here. These experimenters also got similar debriefing aftereffects. Those subjects who were told in Experiment

---

[4] On both the Sociability index and the Interview Performance index, the three-way interaction (Selected Concern × Manipulated Concern × Type of Personality Report) is insignificant as is the two way interaction between selected concern and manipulated concern. All four interaction *F*s are below .20, $df = 1$ and 56. No significant interaction was present for these variables, and it would have been difficult to know what to make of these interactions if they had been significant.

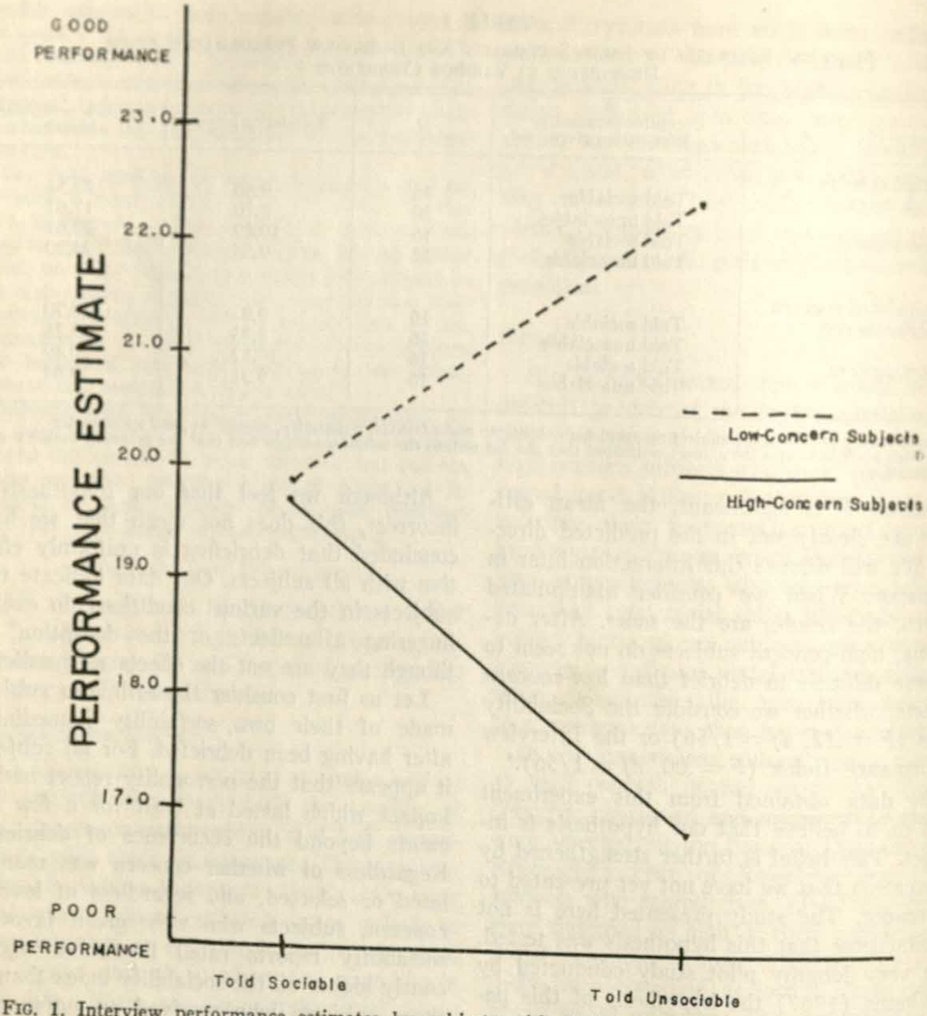


FIG. 1. Interview performance estimates by subjects with high or low fear of rejection scores.

2 that they were immature and uncreative rated themselves as less mature after debriefing ($F = 3.25$, $df = 1/63$, $p = .08$) than did subjects who were told they were mature and creative. The creativity differences were not significant. Since the results of the Abrahams study are so similar to those obtained in this experiment, and since four experimenters were involved in the Abrahams experiment, it is hard for us to believe that our results are simply due to poor debriefing techniques; (*b*) we, like many other experimenters, believe that we make an unusual effort to instruct our students in debriefing procedures. The six experimenters participating in the two studies discussed in this paper were all graduate students who had been trained to debrief subjects and who had conducted several other experiments. Obviously, doing a good job of debriefing was of unusual importance to them in this experiment, and it will be recalled that Experimenter 2 followed a lengthy debriefing procedure.

When we look at the difference in the interview performance estimates between those subjects who received good sociability reports in Experiment 2 and those subjects who received poor reports, we see that the high-sociability-condition subjects do not think that they performed better than do subjects in the low-sociability condition ($F = .22$, $df = 1/56$). Debriefing does not seem to be ineffective here. This interview performance estimate differs in two main ways from the sociability estimate. First, the interview estimates were taken a longer period of time after debriefing than the sociability estimates. Perhaps it takes a few minutes for debriefing to "sink in." Secondly, the Interview Performance Index may measure a different aspect of sociability than the Sociability Index.

In addition to the aftereffects previously discussed, there is one additional finding that seems worth commenting upon; though our manipulated-concern measure seemed to have little impact upon the success or failure of debriefing, there is some evidence that selected concern may be of importance.

From Figure 1 and Table 1 it is clear that a subject's initial degree of concern (with her social abilities) has a marked effect on her postdebriefing estimate of her interview performance.

When we consider only the data from *selected high-concern* subjects, we see that even after the passage of time, debriefing does not seem to be totally effective. Selected high-concern subjects who received good sociability reports in Experiment 2, even after being told these reports were false, estimated that they did better in the interview situation than did subjects who received poor sociability reports. *Selected low-concern* subjects who received good sociability reports estimated their interview performance very much as did high-concern subjects who received the good report. However, the interview performance estimates of the low-concern subjects who received a poor sociability report are markedly different from the estimates of comparable high-concern subjects. These low-concern subjects guessed they did *better* in the interview situation than did subjects in any other group. This finding is peculiar.

In an attempt to explain this result, our first thought was to consider carefully what selected high concern and selected low concern must mean. According to Rosenfeld, the Fear of Rejection test was constructed to parallel the structural form of Mandler and Sarason's Test Anxiety Questionnaire and, presumably, measures the extent to which a subject is concerned with social rejection. From an analysis of our data, however, it appears that in many ways the Fear of Rejection test is a general measure of self-esteem. Subjects classified as low concern by this test tend to say better things about themselves than do high-concern subjects. For example, on both the Sociability Index and on the Interview Performance Index, there is a main effect due to initial level of selected concern of the subject. Subjects low in fear of rejection claim to be more sociable ($F = 6.06$, $df = 1/56$, $p < .05$) and to have done better in the interview situation ($F = 6.20$, $df = 1/56$, $p < .05$) than do subjects high in fear of rejection.

If low concern does in fact reflect high self-esteem, perhaps we can find an explanation for these findings. The apparent failure of debriefing for high-concern subjects is disturbing, but comprehensible. What is peculiar is the high performance estimates made by low-concern subjects who were given low personality results. Perhaps this simply demonstrates that high self-esteem individuals are especially likely to reject unpleasant information about themselves. Cohen (1959), suggested that

> different self-esteem groups are differentially able to fulfill the important acquired motive of maintenance of self-esteem at the highest possible level. Through their use of avoidance defenses . . . the highs are able to maintain a high-level equilibrium . . . whereas the lows are more dependent upon experimental variation [p. 117].

Cohen further noted that the high self-esteem individual may "emphasize enhancing experiences, thereby preserving an insulated but positive self-picture [p. 117]." Perhaps our low-concern (high self-esteem) subjects, when faced with a low personality report, simply enhanced their performance in the interview situation in order to maintain their self-esteem.

This explanation is not entirely satisfactory. If Cohen's formulation is applicable there should be *some* evidence in our experiment that the selected low-concern subjects were "less receptive to experimental variations" than selected high-concern subjects, as Cohen suggests is the case. Low-concern subjects seem, during Experiment 2, to accept the negative personality reports to the same extent as do high-concern subjects. Acceptance of the negative or positive evaluations was measured by the extent to which subjects indicated that they were similar to a sociable or unsociable stimulus girl. Low-concern subjects who were told they had low sociability scores were just as likely to agree that they were similar to a low sociability girl as were high-concern subjects (interaction $F = .03$, $df = 1/56$). Further, even immediately after debriefing, there is no evidence that low-concern subjects, who were told in Experiment 2 that they had low sociability scores, were especially anxious to inflate their estimates of their own sociability (interaction $F = .62$, $df = 1/56$). It was not until they were quizzed several minutes after debriefing about their performance in another situation that low-concern–poor-evaluation subjects overestimated their performance.

We conclude two things from the preceding study:

1. The question of whether or not it is more difficult to successfully debrief concerned subjects than unconcerned subjects of information relevant to their current concerns remains unanswered. We can only say that two lengthy attempts to demonstrate this effect have been unsuccessful. Whether or not a stronger concern manipulation would produce the effect is, of course, a moot question, but we have been unable to produce evidence of even a slight tendency for subjects to behave in the predicted manner.

2. We have presented evidence that debriefing might not be as immediately effective as experimenters have hoped and assumed. This evidence is distressing for a number of reasons. First of all, it is disturbing that in the present experiment and in the Abrahams experiment, even after a very lengthy and thorough debriefing (probably atypical in thoroughness), subjects still behaved to some extent as though the debriefing had not taken place. Subjects behaved in this manner even though they had voiced to the experimenter their understanding that the manipulation was false, their understanding of the true purpose of the experiment, and even though, by their manner and replies, the experimenter had been satisfied that they did indeed understand the nature of the deception.

Even more disturbing is the evidence that the aftereffects of debriefing might be complex, unpredictable, and may depend in part upon the personality traits of the subjects. The nature of the effect of personality traits in the present experiment were not totally explicable to us. Aftereffects in the Abrahams experiment were also present and somewhat inexplicable. The success of debriefing in that experiment was influenced by several significant interactions between sex of subject, sex of experimenter, and treatments. At the time the Abrahams experiment was run, we were willing to conclude that the significant interactions obtained were perhaps due to chance, to experimental error, to measurement error, and so on. The results of the present experiment, however, combined with the results of the Abrahams experiment, have aroused our suspicions and anxiety that there are often residual effects of debriefing, and that these effects appear to be complex and not easily interpreted. What is needed now is the generation of hypotheses and experimentation which would help us predict exactly what effects can be expected, with what kinds of people under what types of conditions, as well as an investigation of the effectiveness of various types of debriefing procedures.

REFERENCES

ABRAHAMS, D. The effect of concern on debriefing, following a deception experiment. Unpublished master's thesis, University of Minnesota, 1967.

COCHRAN, W. G. The comparison of percentages in matched samples. *Biometrika,* 1950, **37**, 256–266.

COHEN, A. R. Some implications of self-esteem for social influence. In I. L. Janis and C. I. Hovland (Eds.), *Personality and persuasibility*. New Haven: Yale University Press, 1959. Pp. 102–137.

KELMAN, H. C. The human use of human subjects: The problem of deception in social-psychological experiments. Paper presented at the meeting of the American Psychological Association, Chicago, 1965.

ORNE, M. T. Demand characteristics and their implications for real life: The importance of quasi controls. Paper presented at the meeting of the American Psychological Association, Chicago, 1965.

ROSENFELD, H. M. Social choice conceived as a level of aspiration. *Journal of Abnormal and Social Psychology,* 1964, **68**, 491–499.

SILVERMAN, I. Motives underlying the behavior of the subject in the psychological experiment. Paper presented at the meeting of the American Psychological Association, Chicago, 1965.

SUTCLIFFE, J. P. A general method of analysis of frequency data for multiple classification designs. *Psychological Bulletin,* 1957, **54**, 134–317.

(Received August 25, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 381–390

PERSONALITY, DEFENSIVENESS, AND EDUCATIONAL ACHIEVEMENT

BINYAMIN Z. LEBOVITS[1] AND ADRIAN M. OSTFELD[2]
Roosevelt University *University of Illinois College of Medicine*

The scores of a group of 1,805 subjects on the MMPI and a sociological questionnaire were examined in order to determine how the MMPI scores vary as a function of education and defensiveness. Specifically, the findings were: (a) the higher the educational level, the greater was the frequency of high K scores, (b) with K scores held constant, high education was associated with higher Mf scores and low Lie- and F-scale scores, (c) Ss with more education more frequently had highpoint scales of Depression, Hysteria, and Paranoia compared with subjects of less education. The results indicate that both defensiveness and educational achievement influence the level of the MMPI scores and the nature of the highpoint scales. They suggest that educational achievement and defensiveness should be controlled whenever groups are compared; that defensiveness should be treated as a personality characteristic and as a response set; and that more sophisticated statistics be explored in making a fuller use of personality tests.

During the last decade the use of objective personality tests has increased appreciably. Of these tests the Minnesota Multiphasic Personality Inventory (MMPI) is one of the more popular (Sundberg, 1961). However, the information presently available in regard to the MMPI and many of the other tests is inadequate. In particular, few systematic studies have been reported on the interaction between socioeconomic status and responses on tests of personality. It appears logical to assume that such an interaction exists and that the existence of socioeconomic classes implies interval differentiation and thus, class-specific response sets.

Interval differentiation refers to the idea that, given the fact that social institutions impose role patterns that differ along many dimensions, the resultant effect is to create groups who are characterized by different patterns of behavior. Sociological studies have suggested that the most important of these dimensions are education, occupation, dwelling area, income, type of home, and associations (Warner, 1960). In both the family of rearing and procreation these factors reflect social forces that pattern behavior. Thus, within any society discrepancies may be expected (subcultural relativity) as one moves from subgroup to subgroup.

Obvious as these relationships may appear, they are seldom treated in the relevant literature. In this paper an attempt is made to elucidate the nature of changes in MMPI code scores and profiles associated with levels of defensiveness and with education. The latter variables have been selected because the literature suggests that they are correlated with patterns of MMPI scores (Dahlstrom & Welsh, 1960, pp. 146–157; Reissman & Miller, 1965; Vernon, 1964; Welsh & Dahlstrom, 1956).

However, it is important to note that our definition of defensiveness is broader than that of other writers, for example, Heilbrun (1961). By defensiveness we mean not only (*a*) the tendency of a person to deny negative characteristics and to admit to positive characteristics, *but also* (*b*) the capacity of a person to respond in an adjustive fashion whenever he is confronted by situations that threaten him, arousing anxiety or guilt. Thus, the term refers to much more than "faking good" and "faking bad." That the *K* scale is associated with personality characteristics and with ego strength has been demonstrated by such studies as those of Heilbrun (1961) and Younge (1966). For a complete description of the personality characteristics ascribed to the *K* scale, the reader is referred to the MMPI manual (Dahlstrom & Welsh, 1960).[3]

---

[1] Also Research Associate in Psychiatry, University of Illinois College of Medicine, Chicago.

[2] Also Head of Preventive Medicine and Community Health, University of Illinois College of Medicine, Chicago.

Specifically, we seek answers to the following questions:

1. Is the distribution of *K* scores significantly different among the various educational groups?
2. How extensive are the changes in the means of the MMPI scales as the level of defensiveness varies?
3. When defensiveness is controlled, what differences are found on the MMPI scales among the various educational levels?
4. Is the distribution of highpoint scales significantly different in the various educational levels? If so, what highpoints are more characteristic of the several educational levels?

METHOD

Procedure

Sample. The material presented here was collected as part of a longitudinal study of coronary heart disease initiated in 1957. The subjects cooperating in the study were employed by the Western Electric Company, Hawthorne Works, in Chicago. The total sample was selected by identifying all male employees between the ages of 40 to 55, inclusively, who had a minimum work record of 2 years, and had no known history of coronary heart disease. There were 5,397 such men. Of these, 3,102 were randomly selected and invited to participate in the study; 2,107 (68%) accepted. Of this number, 117 were later excluded, although they are being examined annually in the same manner as the others. The total initial sample, therefore, numbered 1,990 men.

Most of these subjects were employed in light assembly work, but the sample included some managerial, engineering, and clerical personnel. Approximately three-fourths of the men were second and third generation Americans of German, Polish, and Bohemian stock. The remaining quarter were largely of older American stock of English, Scottish, and Scotch-Irish extraction. Occupation, income, education, house type, dwelling area, and social class varied widely, ranging from upper middle to upper lower.

During the first-year examinations the MMPI was administered to all subjects; during the third annual examination a sociological questionnaire was administered to all the subjects who were still participating in the study. The same person administered the test, and the place of examination was the same for all the subjects.

The sociological questionnaire employed here (the Index of Urban Status) was designed to estimate the social status of each of the subjects—a modification of the *Index of Status Characteristics* (Warner, 1960) constructed by Coleman. It was the result of new data gathered in the Kansas City study by Coleman and associates and was intended for use in midwestern urban environments. The data provided by the questionnaire included place of residence, type of home, occupation, income, extent and place of education, type and amount of participation in voluntary organizations, religion, and place of birth for each subject and his spouse.

Each category was rated on a 7-point scale, with a rating of 1 denoting a characteristic of the upper class and a rating of 7, a characteristic typically associated with membership in the lowest social class. The reliability of the judge (R. Coleman) was established by having him rate, after 2 years, a 10% random sample of the questionnaires. A difference of 1 point arose in from 10% to 29% of the cases due to unreliability of the judge or to clerical errors; and the minimum percentage agreement was 71. In the case of educational status, in 83% of the cases there was perfect agreement, and in 99% of the cases the second rating differed from the first by no more than 1 point.

Ranks on educational level were based on the number of years of schooling and the nature and quality of the school attended.

Table 1 lists the frequency with which the subjects were ranked on various categories and gives the class membership associated with each rank. Note that the median educational level of our subjects was 12 grades. Thus, our group was not a representative one in that respect. The median for the general population is approximately 9 grades for those 50 and above, and the median is approximately 10 grades for those 40 to 44 years of age (United States Bureau of the Census, 1960).

Two groups of subjects were excluded: those who failed to take either test, and also those for whom no scores were available for one of the *basic* socio-

TABLE 1

FREQUENCY OF THE RANKS[a] AND THEIR CLASS MEMBERSHIP[b] ($N = 1885$)

	UU	LU	UM	MM	LM	UL	LL
	1	2	3	4	5	6	7
Neighborhood	002	032	223	607	676	294	050
Occupation	000	021	373	421	1050	019	000
Dwelling	000	068	223	1049	404	131	009
Income	000	023	269	505	1053	032	000
Education	005	187	157	572	611	344	008
Class of origin	049	000	372	125	1000	315	000

[a] Numbers 1–7 represent rank.
[b] U = Upper, L = Lower, M = Middle.

---

[3] No attempt will be made to deal with the question of response sets. The authors tend to agree with Rorer (1965) and Rorer & Goldberg (1965) that the importance of response sets has been highly exaggerated. However, for a contrary view, see Rundquist (1966).

TABLE 2

MEANS AND STANDARD DEVIATIONS ON THE MMPI SCALES FOR THE TOTAL SAMPLE AND FOR THE SAMPLE WHEN SORTED INTO EDUCATIONAL LEVELS

Scale	Total		Educational level									
			2		3		4		5		6	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
L	4.26	2.35	3.83	1.99	3.96	2.24	4.06	2.28	4.43	2.44	4.69	2.42
F	3.53	2.70	2.62	1.97	3.16	2.39	3.31	2.30	3.66	2.68	4.36	3.49
K	58.36	9.70	62.05	8.33	59.61	9.78	59.05	9.74	57.69	9.55	55.93	9.66
Hs	52.50	9.60	49.45	7.61	51.18	8.66	51.62	9.36	53.46	9.86	54.55	10.21
HsK	57.57	10.62	56.56	8.61	57.13	9.89	57.09	10.25	58.04	10.83	58.31	11.78
D	58.17	10.78	56.06	9.55	55.13	10.73	57.72	10.89	58.89	10.81	60.21	10.39
Hy	58.13	9.02	60.38	6.94	58.70	8.22	58.75	8.82	57.74	9.32	56.30	9.66
Pd	51.39	9.51	48.06	8.35	50.63	9.47	50.94	9.44	52.35	9.72	52.71	9.32
PdK	55.72	9.74	54.57	8.97	55.58	9.68	55.62	9.81	56.32	9.80	55.62	9.51
Mf	55.67	8.60	59.68	8.47	57.84	9.15	56.27	8.21	54.87	8.32	52.79	8.07
Pa	51.78	8.06	53.17	6.67	52.51	7.24	52.82	7.72	50.49	8.08	51.14	9.16
Pt	47.84	8.60	44.67	6.71	46.46	7.98	47.57	8.67	48.26	8.65	50.00	8.91
PtK	54.15	9.35	53.42	8.29	53.26	8.17	54.56	9.56	54.01	9.29	54.54	9.75
Sc	46.62	7.39	43.78	4.86	45.64	6.33	45.94	6.89	47.14	7.50	48.83	8.79
ScK	52.90	9.18	52.64	7.89	52.68	8.05	52.64	8.97	52.93	8.98	53.55	10.54
Ma	50.52	9.05	49.27	7.79	50.68	8.16	49.97	9.33	51.21	9.08	50.83	9.43
MaK	51.89	9.58	51.44	8.59	52.28	8.41	51.50	9.99	52.56	9.41	51.39	10.02
Admit	2.60	2.68	1.89	2.10	2.25	2.42	2.47	2.65	2.76	2.72	3.06	2.83
Denial	12.20	3.83	13.86	3.28	12.71	3.82	12.75	3.63	11.80	3.88	10.90	3.79
Anxiety	8.29	6.91	5.68	5.14	7.27	6.25	8.04	6.90	8.65	7.08	9.94	7.17
Repression	16.62	4.69	17.23	4.44	16.30	4.42	16.62	4.76	16.55	4.68	16.62	4.78
Age	47.61	4.44	48.22	4.35	46.74	4.52	46.09	4.45	48.00	4.14	49.47	3.96
N	1852		185		158		567		600		331	

logical variables. After these two groups were excluded, there remained 1,852 subjects.

RESULTS

Education and Scale Scores

IBM cards for each of the 1,852 subjects were sorted according to educational level and five groups were obtained, there being too few subjects in Educational Level 1 ($N = 7$) and Level 7 ($N = 6$) to warrant inclusion as independent groups. Table 2 lists the means and standard deviations of the groups and for the total sample. The results show that there appears to be a relationship between educational level and scores on the MMPI scales. On many of the scales (e.g., *K, Hs, Pd, Mf, Pt, Sc,* Admittance, Denial, and Anxiety), the means show a systematic increase or decrease as educational level changes.

The Significance of the K Scores

The significance of the differences on the *K* scale was ascertained by (*a*) examining the distribution of the *K* scores in the various educational levels, (*b*) determining the correlation between the *K* scales and the other scales, and (*c*) dividing the sample into deciles on the *K* scales and assessing the extent of changes in the means. Table 3 shows the frequency of the cases that fell within any one decile. It can be seen that there was a marked tendency for the subjects within the lower educational levels to obtain lower *K* scores and for the subjects in the upper levels to obtain higher *K* scores. The differences in

TABLE 3

DISTRIBUTION OF *K* SCORES IN THE DIFFERENT EDUCATIONAL LEVELS

K (decile)	Educational level					
	2	3	4	5	6	*N*
40's	14	28	96	122	81	341
50's	49	42	178	203	109	581
60's	87	59	196	176	100	618
70's	34	25	85	86	26	256
	184	154	555	587	316	1796

Note.—There were too few subjects with *K* scores in the 30's and the 80's to include in the distribution ($N = 56$).
[a] $\chi^2 = 51.22$ ($p < .001$).

TABLE 4

CORRELATION OF *K* SCALE AND THE OTHER SCALES OF THE MMPI

	Total sample	Educational level				
		2	3	4	5	6
L	42**	36**	53**	42**	48**	45**
F	−32**	−11	−27**	−31**	−34**	−32**
Hs	−35**	−33**	−38**	−36**	−35**	−27**
HsK	25**	33**	32**	25**	25**	29**
D	−10**	−04	−09	−07	−11*	−04
Hy	43**	46**	49**	39**	40**	44**
Pd	−06	04	−08	00	−09*	−07
PdK	42**	50**	41**	45**	40**	44**
Mf	−09**	−12	−22**	−14**	−12**	−13*
Pa	15**	13	19**	21**	13**	03
Pt	−68**	−61**	−73**	−66**	−70**	−69**
PtK	17**	35**	23**	14**	16**	15**
Sc	−59**	−49**	−64**	−56**	−61**	−58**
ScK	37**	64**	50**	40**	35**	25**
Ma	−36**	−20*	−38**	−33**	−40**	−41**
MaK	−12**	03	−11	−10*	−15**	−18**
N	1852	185	158	567	600	331

Note.—Decimal points omitted.
* $p < .05$.
** $p < .01$.

the distribution are highly significant ($\chi^2 = 51.22$, $p < .001$).

Table 4 lists the correlations of the *K* scale and the other scales in the MMPI. Column 1 shows the correlations when the total sample is considered; the remainder of the table shows the correlations when each educational level is taken separately. The pattern indicates that with the exception of *D* and *Pd*, the *K* scale is highly correlated with the other MMPI scales. This is particularly true of the complaint scales and the *K*-corrected scales.

Table 5 shows the means and standard deviations of the total sample when the sample was grouped into deciles on the *K* scale. The means of the clinical scales show important changes as the mean of the *K* scale is varied. With the exception of only four scales (*D, Pd, Mf,* and *MaK*), the differences are substantial. Typically, the differences between Group 1 (where the mean *K* score is 45.6), and Group 4 (where the mean *K* score is 72.8), are at least one-half standard deviation and sometimes one and one-half standard deviations. When the special scales are examined, the differences are even more impressive. For example, the means of the extreme groups on the anxiety and denial scales are, respectively, 16.18 versus 1.96, and 8.21 versus 16.51.

Education and MMPI Scores with K Held Constant

The subjects in each educational level were classified according to *K* level. This procedure produced 20 groups, since each of the five educational levels were further subdivided into four *K* levels. Comparisons were then

TABLE 5

MEANS AND STANDARD DEVIATIONS OF THE TOTAL SAMPLE WHEN DIVIDED INTO DECILES ON THE *K* SCALE

Scale	40's	*SD*	50's	*SD*	60's	*SD*	70's	*SD*
F	4.78	2.94	3.67	2.61	2.88	2.12	2.72	1.83
K	45.60	3.04	55.20	2.78	63.86	2.49	72.79	2.46
Hs	57.19	10.91	53.70	9.48	50.24	8.24	47.81	6.50
HsK	53.97	12.37	56.87	10.71	58.70	9.42	62.09	7.47
D	60.27	13.33	58.65	10.99	56.86	9.39	57.24	8.67
Hy	52.93	9.82	56.46	8.63	60.32	7.52	64.44	5.93
Pd	52.50	11.26	51.35	10.31	50.25	8.08	52.15	6.90
PdK	50.43	10.52	53.98	9.52	57.38	7.53	63.46	6.60
Mf	56.35	8.99	56.10	9.32	55.37	7.87	54.31	7.16
Pa	49.46	9.08	51.27	8.62	52.01	7.13	54.80	5.48
Pt	56.39	8.43	49.20	7.10	43.82	4.78	40.52	4.03
PtK	52.82	11.94	52.78	9.93	54.08	7.14	58.78	5.29
Sc	53.00	8.32	47.22	6.17	43.59	4.22	41.65	2.52
ScK	48.84	11.58	50.67	8.46	54.10	6.46	60.32	4.61
Ma	54.63	9.71	51.18	9.05	48.63	7.67	46.36	6.73
MaK	52.77	10.80	51.70	10.20	51.30	8.56	51.40	7.56
Lie	2.92	1.79	3.93	2.07	4.83	2.30	5.95	2.32
Admittance	4.06	3.20	3.05	2.61	1.87	2.08	1.23	1.60
Denial	8.21	2.74	11.16	2.78	14.09	2.37	16.51	2.11
Anxiety	16.18	6.22	9.52	5.12	4.57	3.22	1.96	1.92
Repression	13.91	4.31	15.78	4.09	17.77	4.18	20.29	4.15
Age	48.03	4.32	47.54	4.40	47.60	4.47	47.23	4.39
N	341		581		618		256	

TABLE 6

MEANS, *F* VALUES, AND SIGNIFICANCE LEVEL ON THE MMPI SCALES WHEN *K* IS 40 TO 49

Scale	Educational level					Analysis of variance—*F*	*F* departure from linearity
	2	3	4	5	6		
F	2.57	3.79	4.47	5.20	5.21	4.27**	0.90
K	45.50	45.54	45.72	45.42	45.78	0.22	0.28
Hs	53.29	53.32	57.08	58.32	57.56	1.76	0.99
HsK	49.29	49.54	53.75	55.74	54.81	1.83	0.73
D	55.71	55.07	59.94	61.61	61.05	1.88	0.86
Hy	52.14	50.75	53.60	53.43	52.17	0.69	0.92
Pd	46.57	53.54	51.13	53.47	53.22	1.76	1.32
PdK	44.93	51.32	49.21	51.34	51.06	1.71	1.27
Mf	59.00	59.75	57.74	55.86	53.44	4.04**	0.37
Pa	52.14	48.82	49.49	49.26	49.32	0.35	0.38
Pt	52.79	54.79	57.08	56.68	56.02	1.14	1.28
PtK	47.36	50.36	54.17	53.00	52.37	1.40	1.63
Sc	50.21	51.25	52.89	53.60	53.22	0.89	0.44
ScK	44.86	46.32	48.80	49.45	49.38	0.90	0.36
Ma	54.50	54.79	54.80	54.87	53.88	0.15	0.13
MaK	53.29	52.86	53.04	52.93	51.91	0.15	0.08
Lie	2.07	2.18	2.90	2.89	3.41	3.60**	0.78
Admittance	2.86	3.36	4.25	4.17	4.01	1.01	0.96
Denial	9.29	8.50	8.36	8.16	7.90	0.96	0.12
Anxiety	12.50	14.39	16.76	16.26	16.42	2.05	1.69
Repression	13.79	13.54	13.45	14.34	14.15	0.68	0.48
Age	50.50	48.43	46.21	48.34	48.88	6.72**	8.36**
N	14	28	96	122	81		

made among the educational levels holding the *K* level constant. Once the group means were determined for each of the 20 groups, four tables like Table 6 were constructed, one table for each *K* level. A simple analysis of variance and a trend analysis (Lindquist, 1953, Ch. 15) were performed. Table 7 presents a summary of the findings. In the case of five of the scales, the analysis of variance shows significant differences among the educational levels in at least three of the four tables. Subjects with relatively little education achieved higher scores on the *F* and Lie scales. Subjects with a more extensive education achieved a higher score on the *Mf* scale. The *Pa* scale shows a curvilinear relationship to education. Subjects with a moderate amount of education tended to achieve higher scores on the *Pa* scale. They were also typically younger in age.

Profile analysis

A 1620 IBM computer was programmed to rank all but the *Mf* scale and then to list the frequency of the various combinations of the two highest scales. The frequency of the various MMPI profile patterns was then determined for the various educational levels. The computer was programmed to take into account differences of even 1 point and the

TABLE 7

SUMMARY TABLE FOR THE ANALYSIS OF VARIANCE FOR EACH LEVEL OF *K*[a]

Scale	40	50	60	70
F	.01	.05	.05	
K			.05	
Hs			.01	.05
HsK			.01	
D			.01	
Hy		.05		
Pd			.01	.05
PdK			.01	
Mf	.01	.01	.01	.01
Pa		.01	.01	.05
Pt			.01	.05
PtK				
Sc		.05	.01	
ScK		.05	.05	
MaK				
Lie	.01	.01	.01	.01
Admittance			.05	.05
Denial		.01	.01	
Anxiety				
Repression				
Age	.01	.01	.01	.01

[a] Figures indicate significance level of the *F* value derived from a comparison of the mean MMPI scores of the educational levels.

TABLE 8

DISTRIBUTION OF HIGH POINT SCALES ACCORDING TO EDUCATIONAL LEVEL

High point	Educational level						N
	2	3	2 & 3	4	5	6	
1–2	1	6	7	14	20	15	56
1–3	12	7	19	29	38	20	106
1–4	0	1	1	7	8	3	19
1–6	0	1	1	2	2	0	5
1–7	0	1	1	3	6	4	14
1–8	1	0	1	1	4	5	11
1–9	0	1	1	8	5	3	17
2–1	7	6	13	20	42	27	102
2–3	9	3	12	27	27	16	82
2–4	2	0	2	15	17	10	44
2–6	1	2	3	9	7	6	25
2–7	6	5	11	31	26	15	83
2–8	2	3	5	4	5	5	19
2–9	2	5	7	5	11	10	33
3–1	16	15	31	37	28	6	102
3–2	7	6	13	18	19	6	56
3–4	11	6	17	22	12	3	54
3–6	5	4	9	10	4	3	26
3–7	1	0	1	1	1	1	4
3–8	3	0	3	6	4	2	15
3–9	6	2	8	11	3	2	24
4–1	0	0	0	7	6	8	21
4–2	3	3	6	15	11	14	46
4–3	9	5	14	15	14	6	49
4–6	0	2	2	4	4	3	13
4–7	1	1	2	4	2	1	9
4–8	3	0	3	11	11	5	30
4–9	1	3	4	9	11	3	27
6–1	0	0	0	4	1	1	6
6–2	1	1	1	4	3	2	10
6–3	4	1	5	7	1	1	14
6–4	1	1	2	5	1	2	10
6–7	0	1	1	1	0	1	3
6–8	0	0	0	2	0	0	2
6–9	3	1	4	8	2	1	15
7–1	1	1	2	1	2	2	7
7–2	2	0	2	7	5	6	20
7–3	2	0	2	2	1	1	6
7–4	0	2	2	2	3	1	8
7–6	0	0	0	2	2	0	4
7–8	1	1	2	5	3	3	13
7–9	0	0	0	4	1	1	6
8–1	2	0	2	2	3	3	10
8–2	0	0	0	1	2	3	6
8–3	1	0	1	2	3	1	7
8–4	0	0	0	1	2	0	3
8–6	0	1	1	0	2	3	6
8–7	0	0	0	4	3	2	9
8–9	1	0	1	1	3	2	7
9–1	2	0	2	7	10	6	25
9–2	4	2	6	8	19	10	43
9–3	6	5	11	8	12	3	34
9–4	3	6	9	17	20	1	47
9–6	2	2	4	7	5	4	20
9–7	0	0	0	3	3	4	10
9–8	1	1	2	4	11	2	19
N	185	158	343	567	600	331	

distributions listed below do not include ties. Table 8 lists the frequency of the various highpoint scales for the educational levels. In the analysis, Educational Levels 2 and 3 were combined due to the paucity of cases in these levels. The figures for the combined group are presented in Column 3 of Table 8. The statistical analysis of the distributions is limited to the frequency of the patterns that begin with a specific scale. Thus, all the patterns beginning with *Hs* were summed and the total frequency of the "*Hs* patterns" used in the analysis. This means that no statistical test is made to assess differences in the distribution of such patterns as *Hs-Hy*, *Hs-Pd*, etc. The analysis is limited in this fashion for two reasons: (*a*) There is no reason to believe that educational levels should show difference in frequency on all patterns. Too little data is available to determine which of the specific patterns should be singled out for consideration; and (*b*) if all the patterns were considered the investigation could be capitalizing upon chance findings.

TABLE 9

FREQUENCIES AND CHI-SQUARE VALUES FOR THE HIGH POINTS BEGINNING WITH EIGHT OF THE CLINICAL SCALES (*Mf* EXCLUDED) ACCORDING TO EDUCATIONAL LEVEL

Scale	Educational level				χ^2
	2 & 3	4	5	6	
HsK	31	64	83	50	7.82*
D	53	111	135	89	15.04**
Hy	82	105	71	23	47.59***
PdK	31	65	59	40	—
Pa	13	31	12	8	11.96**
PtK	10	23	17	14	—
ScK	5	11	18	14	6.59
MaK	34	54	80	30	6.30

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 9 lists the frequency in the four educational levels with which one of the eight clinical scales is the first one in the various highpoint combinations. In each case the chi-square test was utilized to assess the significance of the differences in frequency. On this basis four scales differ in frequency: the *HsK*, the *D*, the *Hy*, and the *Pa*. Individuals with relatively more education tended not to have highpoint scales beginning with *HsK* and did tend to have highpoint scales beginning with *D*, *Hy*, and *Pa*.

DISCUSSION

Subjects in the various educational levels were compared in regard to their scores on the MMPI. It was found that the distribution of *K* scores in these groups was significantly different and that, in part, this accounts for the differences on the other MMPI scales. The importance of the *K* scale is demonstrated by the highly significant correlations between *K* and the other MMPI scales, and the changes in mean scores as the deciles of the *K* scores are varied. Although age is a variable that also influences the patterning of the MMPI scales, it was postulated that in view of the narrow age range of the subjects included in the study, age was of lesser importance. Support for this view is provided by the examination of the correlations between age and the other MMPI scales. The highest correlation is between *HsK* and age (r is .18, p is .01). Although highly significant, it can account for only a small proportion of the variation among the scores. When the educational levels are compared holding *K* level constant (i.e., within 10 points), five factors appear to be consistently different, that is, appear as significant factors on at least three of the four levels examined. These are the *F*, the Lie, the *Mf*, and the *Pa* scales, and age. The more educated subjects achieved lower scores on the Lie and *F* scales and higher scores on the *Mf* scale. On the *Pa* scale, the subjects in the middle range of educational achievement scored higher. And finally, age shows a curvilinear relation to educational level. Those with the most and least education (Educational Levels 2 and 6) were older than those with an intermediate amount of education.

The MMPI is typically interpreted on the basis of the patterning of the scales, that is, highpoint scales are the fundamental anchors for the process of interpretation. The scales were, therefore, ranked and the distribution of the highpoints examined. Differences were found as a function of educational level. Subjects with greater amounts of education less frequently had highpoint scales beginning with *HsK*, but tended more frequently to have highpoint scales beginning with *D*, *Hy*, and *Pa*.

K scale. One of the fundamental findings of this study is the difference in the distribution of *K* scores in the various educational levels. Subjects with more education achieved significantly higher scores on the *K* scales than those with lesser amounts of education. However, the *K* scale is interpreted in two ways: (*a*) as a reflection of test-taking attitudes and, thus, as an indicator of the extent to which corrections must be made, and (*b*) as a measure of specific personality variables, in particular, defensiveness. As presently constituted, these two aspects of the *K* scale cannot be easily isolated. There is some suggestion that even when factor analysis is used, these two facets of the *K* scale may not be discriminable. It remains a moot question, then, as to the meaning of a high *K* score: it may mean that the subject is merely censuring his answers carefully and/or that the subject is characterized by a certain set of traits. It is likely, however, that a high *K* score carries both of these meanings and such an inference is made by the authors.

The differences in the distribution of *K* scores suggest that whenever groups with different distributions of education are compared, differences are likely to emerge on a number of MMPI scales. These differences may reflect the greater defensiveness (i.e., guardedness and ego strength) of the more educated subjects, and the lesser sophistication and ego strength of the more poorly educated subjects. On the one hand, the more educated subjects probably possess greater insight into and knowledge about the significance of the MMPI statements and their answers to them. They may tend, therefore, to be more cautious about admitting to characteristics and attitudes that imply abnormality. This may reflect the middle-class concern with the projection of the right kind of image as opposed to the lower-class concern with *being* the right kind of person (Kohn, 1959). Thus, in regard to some areas, for example the kind of items that appear on the Lie scale, the more educated subjects do not find it necessary to project an image of goodness. They recognize that such behavior (e.g., "I always read the editorials in the newspaper every day.") is based on standards that are impossibly high. On the other hand, the higher *K* scores indicate the ability to minimize traumatic material—the anxiety or guilt-provoking aspects—of experiences, thought, and actions.

Educational level and the clinical scales. When comparisons are made among the educational levels holding *K* constant (i.e., within 10 points), four scales are found to differentiate consistently among the various educational levels, that is, at least on three of the four levels of *K*. Individuals with more education scored lower on the Lie and *F* scales and higher on the *Mf* scale. Subjects with moderate amounts of education achieved higher *Pa* scores and typically were younger in age. When these patterns are interpreted using the MMPI Handbook (Dahlstrom & Welsh, 1960) as the source of hypotheses, the following emerges. The more educated subjects appear to have felt less dissatisfied with themselves and view themselves as being socially rooted. On the one hand, their views of themselves and their relationship to society are characterized by a greater amount of introspectiveness and self-awareness. On the other hand, the self-acceptance and social rootedness are not necessarily a function of greater sociability. In fact, the *Ma* scale is one of the few scales that does not show differences in means under any of the conditions examined in this paper. Thus, a great deal seems to hinge upon the subject's ability to interpret his way through difficulties. It is in this way that the *Mf* can be interpreted. The *Mf* scale reflects the greater tendency of the more educated subjects to be introspective, to engage in thought and fantasy, and to manifest a sensitivity that carries with it some connotation of ineptness. While the elevation of the *Mf* scale in men is a measure of the presence of the more effeminate patterns of our culture (e.g., interest in music, arts, etc.), it also implies some feeling of awkwardness insofar as a task orientation is concerned. Action is cautiously approached and cautiously explored before a commitment is made. In part, this may be a function of a greater awareness of limitations and a greater need for social approval, and in part, due to the fact that an occupationally and geographically stable group of subjects are the only ones included in this study.

The *Pa* scale was found to be higher among the subjects with a moderate amount of education. The low *Pa* scores reflect two patterns: a tendency for the more educated subjects to be more distant and more secure; and for the least educated subjects to be more direct, open, and resigned. The more educated subjects occupy positions of authority and must render decisions about the performance and potential of their subordinates. These tasks make it more difficult for them to assume a completely empathic attitude and to establish very close personal relationships with their subordinates. In our technologically oriented society with its emphasis on education, individuals with a great deal of education can rest secure in the knowledge that they possess skills that make them indispensible. Subjects in the lower educational levels tend to be more blunt and direct. Their pattern of discourse and modes of relationships are less characterized by interpersonal and verbal maneuvering. They say what they feel. These individuals also, middle aged, have come to recognize their educational and occupational limitations and have resigned themselves to their present station. However, the individuals with middle levels of education have neither the security of educational achievement nor the willingness to write off advancement. They are most concerned about the opinions of others and the degree to which they are being treated fairly. Thus, in this context, the *Pa* scores are taken to reflect the presence or absence of interpersonal closeness and security, the acceptance of one's educational limitations, and concern over one's reputation and the objectivity of the immediate authority figures in one's life.

Highpoint scales. The more educated subjects are more likely to have profiles beginning with *D*, *Hy*, and *Pa*. Subjects with lesser amounts of education are more likely to have profiles beginning with *HsK*. Using Pressey and Kuhlen's (1957, pp. 292–297) description of the middle years, the following interpretation emerges. The educated subject tends to reach the apogee of his career toward the middle years of his life. Around this time, he becomes aware of the degree to which he can attain the goals he has set for himself. Thus, he is more fully aware both of his limitations and of the potential frustrations and rewards of the work world. As his opportunities begin to wane, the educated individual faces the need to restructure his goals, to accept his oncoming decline, and to prepare for a supervisory role rather than an action role. It is likely that the three types of profile patterns reflect different approaches to these problems. The individuals who possess patterns beginning with the *D* scale are characterized by a tendency toward self-depreciation and a resultant feeling of inadequacy. These individuals seek to inspire and to elicit support from their social environment. They appear worried, unhappy, and manifest a great deal of tension. The second pattern may be characteristic of individuals who respond by increased attempts to actualize their abilities, realize their goals, and to disregard physical change. Their behavior is marked by energetic, enthusiastic application of skills and an imaginative exploitation of whatever opportunities arise. Finally, the third pattern may be found in individuals who become highly ego-involved and identify themselves with their work. These may be the individuals who can accept a supervisory role and focus on the application of their past experience—the imaginative exploitation of their experience. Thus, they apply themselves both to the personal-social contexts of their activities and the physical requirements of their tasks.

Individuals with little education tend to reach the apogee of their work career in the latter thirties and early forties. For them, the task in the remaining productive years of their life is to reconcile themselves to the fact that they are now in a work plateau. Aging can only lead to debilitation and the roll of the fifth wheel. These subjects must learn both to disregard the fact that opportunities for advancement have ceased for them, and to accept the fact that their meager education ill prepares them to face the complexities of modern living. Thus, attention tends to become focused on the status of the body. To be healthy is important. With little education what else does an individual have to sell other than the services his body can perform? Moreover, ill health is one way of obtaining comfort and support. Not only is the doctor one of the few people whose role prohibits him from rejecting inferior (sick) people, but illness is one of the few grounds upon which one can accept help and comfort.

Thus, the data suggest that studies utilizing the MMPI may be more fruitful if the characteristics of the *K* scale and scales like the *Mf* are taken into consideration. The *K* score has an important effect on the nature of the MMPI pattern. The *K* score reflects both test-taking attitudes and personality characteristics. As such, it is unwise to control for differences on *K*. But unless the differences on *K* are controlled, differences may automatically be found on other scales. The use of noncorrected scores does not obviate this difficulty. The *K* scores are highly correlated with the noncorrected scores as well as the corrected scores. More sophisticated techniques of analysis may be the only solution.

The *Mf* scores are highly constant. That is, differences among the mean *Mf* scores emerge whenever educational level is allowed to vary. This is true whether or not the *K* scores are held constant. Special attention needs to be paid to this characteristic of the *Mf* scale, for example, in actuarial studies (Marks & Seeman, 1963). If the *Mf* is used to derive highpoint scales, interpretations need to be qualified as a function of educational level.

However, before conclusive recommendations can be made, other populations must be studied. Our sample is a special one. One-third of the subjects requested to participate in the study refused to do so. Only subjects who had been at Western Electric for 2 years were considered for the study. The population consists largely of individuals who are second and third generation Americans of German, Irish, and Bohemian stock. The ages of the subjects fall into a narrow range and the educational level of the subjects is higher than that of the general population.

REFERENCES

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook.* Minneapolis: University of Minnesota Press, 1960.

HEILBRUN, A. B. Psychological significance of the MMPI K scale in a normal population. *Journal of Consulting Psychology,* 1961, **25**, 486–491.

KOHN, M. L. Social class and parental values. *American Journal of Sociology,* 1959, **64**, 337–351.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MARKS, P. A., & SEEMAN, W. *Actuarial description of abnormal personality.* Baltimore: Williams & Wilkins, 1963.

PRESSEY, S. L., & KUHLEN, R. G. *Psychological development through the life span.* New York: Harper, 1957.

REISSMAN, F., & MILLER, S. M. Social class and projective techniques. In B. L. Murstein (Ed.), *Handbook of projective techniques.* New York: Basic Books, 1965. Pp. 95–106.

RORER, L. G. The great response style myth. *Psychological Review,* 1965, **63**, 129–156.

RORER, L. G., & GOLDBERG, L. R. Acquiescence in the MMPI. *Education and Psychology Measurement,* 1965, **25**, 801–817.

RUNDQUIST, E. A. Item and response characteristics in attitude and personality. *Psychological Bulletin,* 1966, **66**, 166–177.

SUNDBERG, N. D. The practice of psychological testing in clinical services in the United States. *American Psychologist,* 1961, **16**, 79–83.

UNITED STATES BUREAU OF THE CENSUS. *Historical statistics of the U. S., colonial times to 1957.* Washington, D. C.: United States Bureau of the Census, 1960.

VERNON, P. E. *Personality assessment.* New York: Wiley, 1964.

WARNER, W. L. *Social class in America.* New York: Harper & Row, 1960.

WELSH, G. S., & DAHLSTROM, W. G. *Basic readings on the MMPI in psychology and medicine.* Minneapolis: University of Minnesota Press, 1956.

YOUNGE, G. Certain consequences of applying the K factor to MMPI scores. *Education and Psychology Measurement,* 1966, **26**, 887–893.

(Received July 30, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 391–399

GENERALIZATION AND AWARENESS IN VERBAL OPERANT CONDITIONING[1]

FALAK THAVER AND WILLIAM F. OAKES
University of Hawaii

Students from introductory psychology classes at the University of Hawaii initially participated in a verbal conditioning session. Following the conditioning task Ss performed a generalization task, at the end of which they were given 2 postexperimental interviews for the assessment of awareness. It was found that awareness of the response-reinforcement contingency on the conditioning task was not related to amount of generalization. Awareness of the influence of the initial conditioning task or of the relationship between the 2 tasks while performing the generalization task had no effect on generalization. Further, intertrial activity did not produce any difference in either conditioning or generalization. Results were interpreted as supporting the reinforcement interpretation of verbal conditioning.

The present study was designed to investigate the generalization to a different situation of a response reinforced through verbal operant conditioning (VOC) procedures and the relationship between such generalization and the subject's awareness. The questions of generalization and awareness, both being of practical as well as theoretical importance, are currently ambiguous in VOC. Although the concepts have been investigated separately, their interrelationships have not been investigated heretofore with respect to VOC.

A number of investigators (Dinoff, Horner, Kurpiewski, Rickard, & Timmons, 1960; Moos, 1963; Rogers, 1960; Weide, 1959; Wimsatt & Vestre, 1963) have failed to find generalization in previous VOC studies. However, positive evidence for generalization has been found by several investigators. Timmons (1959) reinforced responses related to "buildings" and found that the subsequent drawings of such subjects were related to buildings. Greenspoon and Thompson (1959) reinforced animal responses and found more of them later on Card X of the Rorschach. Deering (1958) reinforced angry or egotistical responses and found evidence of the effect on a posttest involving subjects' completion of incomplete sentences referring to social situations. Krumboltz and Thoresen (1964) reinforced verbal information-seeking behavior in a counseling-type situation and afterward found such subjects engaged in more overt information-seeking behavior, such as writing to colleges, etc. Oakes, Droge, and August (1961), in a group discussion, reinforced statements consistent with particular alternative solutions to a problem situation and later found such alternatives selected predominantly by such subjects in a free-choice situation. Insko (1965) reinforced responses favorable or unfavorable to a local Hawaiian festival in telephone interviews and found that the attitudes of subjects toward the festival on a local issues questionnaire administered later in a classroom situation had been influenced in the reinforced direction. Singer (1961) reinforced prodemocratic responses to items on one scale and found this generalized to a related scale.

Thus we find a number of studies which have succeeded in demonstrating generalization and a number which have failed to do so. As yet, sufficient research has not been conducted to permit us to specify the conditions under which one is likely to find generalization and those under which one is not likely to do so.

The role of awareness in VOC is also somewhat ambiguous at present. There are essentially two points of view on the matter. Those holding the cognitive position (e.g.,

---

[1] Data reported herein were obtained by the first author in connection with her doctoral dissertation under the direction of the second author. This report was prepared by the second author. The authors express their appreciation for the helpful suggestions of Chester Insko during the planning of the research.

Dulany, 1962; Spielberger, 1962) maintain that the increased emission of members of the reinforced response class in VOC is dependent on subjects' awareness of the response-reinforcement contingency. The commonly found positive correlation between degree of such awareness (based on retrospective reports) and amount of increased emission of members of the reinforced response class is presented as evidence for the cognitive interpretation of VOC. Those holding the reinforcement position (e.g., Dixon & Oakes, 1965; Krasner, 1962; Oakes, 1967; Verplanck, 1962) maintain that although a correlation between awareness and conditioning is usually found in VOC, the effect of the reinforcement is relatively direct, and not *necessarily* mediated by awareness of the response-reinforcement contingency. Evidence presented in support of the reinforcement interpretation of VOC includes the finding that reported awareness is a function of a number of experimental variables independent of the amount of conditioning (Krasner, 1962), and that interfering with subjects' opportunity to formulate and test hypotheses regarding the response-reinforcement contingency during acquisition eliminates the awareness-conditioning relationship usually found, without eliminating the conditioning (Dixon & Oakes, 1965; Oakes, 1967).

The results of the present study bear on the awareness question in that, if the cognitive interpretation is correct, one would expect to find no conditioning and hence (a fortiori) no generalization without associated awareness on the part of the subjects, whereas, if the reinforcement interpretation is correct, one might (depending on experimental conditions) find such conditioning and generalization without its being related to awareness.

The design of the present study involves a standard VOC procedure, in which for half the subjects an intertrial activity interferes with their opportunity to formulate and test hypotheses about a response-reinforcement contingency. The VOC procedure is followed by a test of generalization of the effect of the reinforcement to a different task in a different situation. The generalization task is followed by assessment of awareness. We are attempting to determine whether the effect of the reinforcement will be evident in the performance of subjects on the VOC task, whether it will be evident in the generalization task, whether it will be influenced by the intertrial activity, and how all these variables will be related to awareness.

METHOD

Subjects

Subjects were 96 students from introductory psychology courses at the University of Hawaii who got additional academic credit for volunteering to participate in such outside experiments.

Apparatus

A sentence-construction task similar to that devised by Taffel (1955) was employed as the operant-conditioning procedure. The stimulus materials for the conditioning task consisted of 75 3 × 5-inch index cards of assorted colors. A hostile and neutral verb were typed on each card. Hostile verbs were taken from the list used by Buss and Durkee (1958). As the neutral verbs provided by Buss and Durkee were not sufficient, additional ones were added to the original list in a manner similar to theirs. Also, on each card colored symbols (square, triangle, circle, and cross) were drawn at the corners in a random order. Verbs placed on each of the cards were matched for their frequency on the basis of the Thorndike-Lorge (1944) word count. The order of placement of the verbs on the cards and the color of the index cards used were randomized for the series of 75 trials.

A screen was used for the presentation of the cards, with an opening in the center of the panel about the size of the index card. A slot beneath the space through which the cards were shown was made in order to fit the memory drum. Colors (26 shades in all) were displayed with the memory drum at the rate of one per second. A stopwatch was used for timing the intertrial interval of 8 seconds.

Pictures 3 and 5 of the TAT were used for the generalization task. There were two interviews used at the end of the experiment. Interview 1, which dealt with the conditioning session, was a modification of the one used by Spielberger (1962) to fit the present procedure. Interview 2 dealt with questions pertaining to the generalization task and was constructed by the writers. Both the interviews are reproduced in Table 1.

Experimental Design

There were eight groups of subjects. Half the subjects were reinforced for making sentences with hostile verbs and the other half for using neutral verbs. In an orthogonal design, half the subjects also had the intertrial activity (color naming), and the other half had no such intertrial activity. At

TABLE 1

INTERVIEWS 1 AND 2

I would like to ask you some questions about the experiment that you just participated in. In answering these questions, it is important for you to think back to when you were going through the cards.

1

Did you usually give the first sentence that came to your mind?

How did you go about deciding which of the two verbs to use?

Did you think you were using certain kinds of verbs more often than others? Which kind? Why?

What do you think the purpose of this experiment was?

While going through the cards, did you think you were supposed to make your sentences in any particular way or that you had to change the way in which you made your sentences? In what way?

Did you notice anything about the experimenter while going through the cards?

Do you remember her saying "good" occasionally?

What did her saying "good" mean to you?

Did you try to figure out what made her say "good" or why or when she was saying "good"?

Did you think there was any purpose to her saying "good"?

While going through the cards, did you think her saying "good" had anything to do with some special types of verbs you used to make your sentences with?

There were about 75 trials in the experiment, when did this idea occur to you?

Were you aware of its influence on your selection of verbs?

Did you at any time during the course of the experiment consciously decide that, since the experimenter was saying "good" in respect to some particular class of verbs, you were going to try and make all your sentences with those class of verbs?

2

How did you go about making stories to the pictures you saw?

Did you enjoy writing the stories?

Did your participation in the first experiment influence you in any way as you were writing the stories? In what way?

While you were writing the stories, did you think that these two experiments were related? How?

How did this recognition affect the content of your stories? Please indicate specifically.

the end of the generalization task, which immediately followed the conditioning task, subjects were given two postexperimental interviews for the assessment of awareness. Again, half the subjects were given Interview 1 prior to Interview 2, and the other half were given Interview 2 prior to Interview 1.

Procedure

Subjects were seen individually by the experimenter. Before beginning the experiment proper, the experimenter initiated a casual conversation to establish rapport. For the conditioning session, subjects were then instructed to make up sentences using one of the two verbs on the cards that were shown to them at intervals of 5 seconds. The first 15 trials were used to establish the operant level. Starting with Trial 16, the experimenter said, "good," each time a hostile (or neutral) verb was used. Prior to Trial 16, the experimenter said nothing in response to subjects' sentences. For subjects with the intertrial activity, colors were displayed through the memory drum, and subjects were instructed to name them as accurately as they could.

At the completion of the conditioning task, each subject was informed that the experiment could not be continued for a short time as the experimenter had to attend to some computations, but that during the time the experimenter was away the subject could participate in another short 15-minute experiment and earn some additional credits.

This "second experiment" was actually the generalization task. Subjects were taken to another room and were asked to follow instructions they found in the folder lying on top of the table. Subjects were told that this second experiment was being conducted by the chairman of the psychology department who was collecting data for some personality research. Subjects were instructed to write a story at least one page long for each of the two TAT cards in the folder.

At the end of the 15 minutes the experimenter returned and led the subject back to the room where the conditioning part of the experiment had taken place. The postexperimental interviews were then conducted.

At the conclusion of the experiment the subject was urged not to discuss any aspect of the experiment with his friends or acquaintances until after a date when all the data would have been collected.

RESULTS

Conditioning Task

The first analysis was made using data from all subjects to determine whether the reinforcement had an effect. The number of sentences using hostile verbs per subject per block of trials was taken as the dependent variable, and an analysis of covariance was made with the value of the dependent vari-

TABLE 2

SUMMARY OF ANALYSIS OF COVARIANCE OF NUMBER OF HOSTILE VERBS USED DURING REINFORCED BLOCKS OF TRIALS WITH NUMBER USED ON OPERANT (UNREINFORCED) TRIALS AS COVARIATE

Source of variance	*df*	Adjusted *MS*	*F*
Between *S*s	94		
Direction of reinforcement (A)	1	346.285	35.785**
Intertrial activity (B)	1	.645	—
A × B	1	3.518	—
Error	91	9.674	—
Within *S*s	288		
Blocks of trials (C)	3	29.621	11.829**
A × C	3	46.472	18.559**
B × C	3	2.694	1.07
A × B × C	3	—	—
Error	276	2.504	—

** $p < .01$.

able on the operant trials as a covariate. A mixed-model design resulted, with the between-subjects treatments being A (response class reinforced) and B (intertrial activity), and the within-subjects treatment being C (blocks of reinforced trials). Table 2 presents a summary of that analysis.

It can be seen from Table 2 that there were significant main effects of A and C, and a significant A × C interaction. Thus the subjects reinforced for use of hostile verbs used more hostile verbs than those reinforced for the use of neutral verbs (A effect). Over blocks of trials, the overall number of hostile verbs used declined (C effect); that is, the neutral verbs conditioned better than the hostile verbs, although the neutral and hostile verb groups differed throughout. The A × C interaction is the result of the increasing difference over trials of the neutral and hostile verb groups.

The lack of a significant B effect or a significant interaction involving B indicates that the intertrial activity had no effect on conditioning over all subjects.

The second analysis made concerned the effect of subjects' awareness of a response-reinforcement contingency on performance on the conditioning task. The interview protocols for all subjects were given to two judges (advanced graduate students who had previously conducted VOC research) for independent ratings of awareness. The agreement of the judges was 85% with respect to awareness. Twenty-three of the 96 subjects were classified as aware by the judges, 14 from the hostile groups and 9 from the neutral. Twelve of these had the color naming and 11 did not. These "aware" subjects were subjects who may have reported the occurrence but neither the significance nor the distribution of the contingent stimulus.

An unweighted-means analysis of covarience was then made, again with operant values as a covariate, but with the additional factor of awareness. The factors in the covariance analysis were thus A (response class reinforced), B (aware versus unaware), C (intertrial activity), and D (blocks of trials). Table 3 presents the summary of that analysis.

It can be seen from Table 3 that the significant effects were A, D, A × D, A × B, and A × B × D. The interpretation of the effects of A, D, and A × D is the same as in the previous analysis discussed above. The significant A × B interaction indicates a

TABLE 3

UNWEIGHTED-MEANS ANALYSIS OF COVARIANCE OF NUMBER OF HOSTILE VERBS USED ON REINFORCED TRIALS WITH NUMBER USED ON OPERANT (UNREINFORCED) TRIALS AS COVARIATE, CONSIDERING SUBJECTS' AWARENESS

Source of variance	*df*	Adjusted *MS*	*F*
Between *S*s	94		
Direction of reinforcement (A)	1	234.494	26.098***
Aware vs. unaware (B)	1	2.632	—
Intertrial activity (C)	1	1.864	—
A × B	1	36.208	4.029
A × C	1	.545	—
B × C	1	24.154	2.688
A × B × C	1	.149	—
Error	87	8.985	
Within *S*s	288		
Blocks of trials (D)	3	20.816	8.630**
A × D	3	52.623	21.817***
B × D	3	.185	—
C × D	3	2.431	1.007
A × B × D	3	10.293	4.267*
A × C × D	3	2.114	—
B × C × D	3	1.609	—
A × B × C × D	3	2.246	—
Error	264	2.412	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

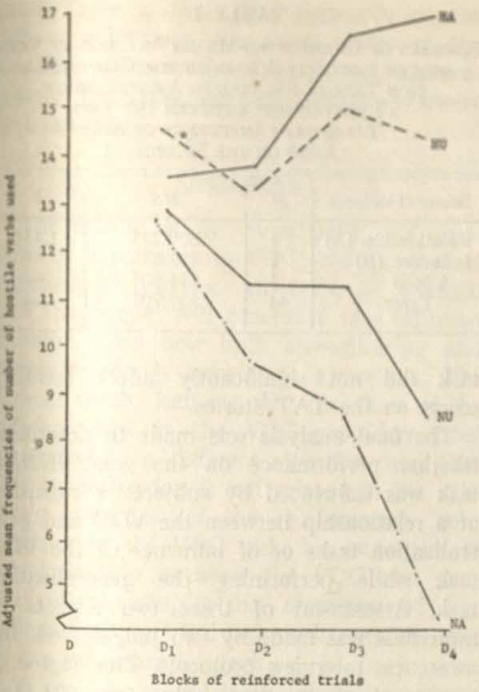


FIG. 1. Use of hostile verbs as a function of trial blocks of hostile aware (HA), hostile unaware (HU), neutral aware (NA), and neutral unaware (NU) groups.

greater effect of the reinforcement for the aware subjects. The A × B × D interaction indicates an increasingly greater effect of the reinforcement for the aware subjects over trials, as can be seen in Figure 1, which represents this interaction graphically.

As in the previous analysis, the intertrial activity variable (C) did not exert a significant effect.

Although the curve for the hostile unaware group in Figure 1 shows little rise, and might therefore suggest the absence of a reinforcement effect, we do have evidence from a previous study in which an unreinforced control group was used (Sarason & Campbell, 1962) which suggests that without reinforcement the rate for hostile verbs declines in this situation. Thus, although within the present study the question of the conditioning of the hostile unaware group per se cannot be answered without an unreinforced control group, and, as a matter of fact, is not crucial for the purposes of the present study, we would suggest that there is a reinforcement effect in the hostile unaware group. Clearly, however, there was a reinforcement effect differentiating the hostile unaware and neutral unaware groups as shown in the following analysis.

The next analysis was made to determine whether there was conditioning of the unaware subjects. An unweighted-means analysis of covariance was made involving only the unaware subjects. Since intertrial activity had not proven significant in any of the preceding analyses, it was excluded from the present one. Table 4 presents a summary of the analysis.

It can be seen from Table 4 that the unaware subjects showed a significant effect of the reinforcement which increased over trials, indicated by the significant A and A × B effects. This reinforcement effect for the unaware subjects can also be seen in Figure 1.

Generalization Task

A list was made of all the verbs used by subjects in the TAT stories written as the generalization task. These verbs were rated on a 5-point neutrality-hostility continuum by two psychology graduate-student judges, with a score of 1 indicating the least hostility, and a score of 5 the most. The correlation of the ratings of the two judges was .70. Each verb used was then given a hostility score, which was the sum of the hostility scores given the verb by the two raters. The dependent variable for the generalization task was then the sum of the

TABLE 4

UNWEIGHTED-MEANS ANALYSIS OF COVARIANCE OF NUMBER OF HOSTILE VERBS USED BY UNAWARE SUBJECTS ON REINFORCED TRIALS, WITH NUMBER USED ON OPERANT (UNREINFORCED) TRIALS AS COVARIATE

Source of variance	*df*	Adjusted *MS*	*F*
Between *Ss*	71		
Direction of reinforcement (A)	1	132.863	18.98***
Error	70	6.999	
Within *Ss*^{a}	226		
Blocks of trials (B)	3	27.601	11.353**
A × B	3	20.957	8.620**
Error	220	2.431	

** $p < .01$.
*** $p < .001$.

TABLE 5

SUMMARY OF ANALYSIS OF VARIANCE OF HOSTILITY SCORES ON THE GENERALIZATION TASK FOR ALL SUBJECTS

Source of variance	*df*	*MS*	*F*
Response class reinforced (A)	1	5148.01	11.23**
Intertrial activity (B)	1	142.59	—
A × B	1	78.84	—
Error	92	458.29	—

** $p < .01$.

hostility scores for the two TAT stories for each subject.

The first analysis of hostility scores on the generalization task was made using scores for all subjects, not considering awareness. The analysis was made on the basis of the conditions of subjects on the VOC task. Table 5 presents the summary of that analysis.

As can be seen from Table 5, the only significant effect was A, the response class reinforced, with subjects reinforced for hostile verbs in the VOC task having higher hostility scores on their TAT stories than those reinforced for neutral verbs. As with the VOC task, intertrial activity had no significant effect on the generalization task.

The next analysis was made to determine whether there was a difference on the generalization task between subjects who were aware or unaware on the VOC task. Table 6 presents the summary of an unweighted-means analysis of variance with the awareness variable added.

It can be seen that awareness on the VOC task did not significantly affect hostility scores on the TAT stories.

TABLE 6

SUMMARY OF UNWEIGHTED-MEANS ANALYSIS OF VARIANCE OF AWARENESS ON THE CONDITIONING TASK AND ITS EFFECT ON HOSTILITY SCORES ON THE GENERALIZATION TASK

Source of variance	*df*	*MS*	*F*
Aware vs. unaware (A)	1	452.64	
Intertrial activity (B)	1	.31	
Class of response (C)	1	4160.95	8.82**
A × B	1	251.53	
A × C	1	16.17	
B × C	1	6.10	
A × B × C	1	120.53	
Error	88	471.63	

** $p < .01$.

TABLE 7

SUMMARY OF UNWEIGHTED-MEANS ANALYSIS OF VARIANCE OF HOSTILITY SCORES ON THE GENERALIZATION TASK AS RELATED TO AWARENESS OF A RELATIONSHIP BETWEEN THE TWO TASKS OR OF INFLUENCE OF THE FIRST ON THE SECOND

Source of variance	*df*	*MS*	*F*
Relationship (A)	1	800.253	1.434
Influence (B)	1	493.838	—
A × B	1	271.805	—
Error	44	557.701	—

The final analysis was made to determine whether performance on the generalization task was influenced by subjects' recognition of a relationship between the VOC and generalization tasks or of influence of the VOC task while performing the generalization task. Assessment of these two aspects of awareness was made by two judges from the awareness interview protocols. The degree of agreement of the two judges was 93 and 96%, respectively.

It was possible to include in this analysis only subjects who had been in the hostile condition on the VOC task, as there were very few subjects in the neutral group who were aware of either any relationship between the two "experiments" or any influence of the VOC task on the generalization task. Table 7 presents the summary of that analysis.

It can be seen from Table 7 that the hostility scores of the TAT stories for the subjects in the hostile group on the VOC task were not influenced by either of these types of awareness.

Table 8 presents the mean hostility scores of the TAT stories for the various awareness and response-class groups. It can be seen

TABLE 8

MEAN HOSTILITY SCORES ON TAT STORIES OF SUBJECTS IN HOSTILE AND NEUTRAL GROUPS WHO WERE AWARE AND UNAWARE

	Hostile	Neutral	Σ
Aware	84.6	68.7	78.4
Unaware	88.9	74.0	80.9
Σ	87.6	73.0	80.3

from Table 8 that the magnitude of the difference between the mean scores of the hostile and neutral subjects was approximately the same whether they were aware or unaware.

DISCUSSION

The present study investigated two questions of importance in VOC: the question of whether a response strengthened through VOC techniques will generalize to a different situation, and how such strengthening and generalization is related to awareness. The present results indicate that the effect of the reinforcement in the VOC procedure does generalize to a different situation, and that this generalization is not related to the subject's awareness of the reinforcement contingency in the VOC task, his awareness of a relationship between the VOC and generalization tasks, or his awareness of any influence of the VOC task on his performance in the generalization task.

It was also found in the present study, as was found by Dixon and Oakes (1965), that the intertrial activity did not interfere with VOC. The Dixon and Oakes study found, however, that the intertrial activity interfered with the frequently found correlation of conditioning and awareness. The present results do not permit us to answer the question of why the present study produced no such finding. A comparison of the procedure in the two studies does suggest some possible explanations, however. Essentially, the intertrial activity of color naming is an attempt at obfuscation. The experimenter attempts to keep the subject so busy between trials and to keep him concerned with so many things that the subject does not have much opportunity to think about his own behavior in relation to the variables that are crucial to the experimenter, that is, the pronoun the subject began the sentence with and the experimenter's reinforcer. In the present study the stubject has much more difficulty identifying the contingency than in the Dixon and Oakes study, whether or not he has the intertrial activity, as additional obfuscating variables were added in the present study. The index cards were different colors, and each card had a symbol in the corner. In addition, the response class of hostile versus neutral verbs is undoubtedly more difficult for the subject to identify than the "I-we" response class reinforced in the Dixon and Oakes study. So, in effect, in the present study, all subjects had obfuscation built into the procedure, but, presumably, the subjects with intertrial activity had somewhat more of it. One possible explanation of the lack of significant effect of intertrial activity on the relationship between awareness and VOC in the present study, then, is that the color naming did not add sufficiently more obfuscation to the non-color-naming procedure to produce such an independent effect. Another contributing factor to this difference in results as compared with the Dixon and Oakes study is based on differences in the interviews used to assess awareness. Dixon and Oakes used the Dulany (1962) interview, whereas the present study used a modification of the Spielberger (1962) interview.

The present results are consistent with the results of Dixon and Oakes and Oakes (1967) in supporting the reinforcement interpretation of VOC findings rather than the cognitive interpretation.

Though the cognitive theorists have thus far restricted their investigation to the simple conditioning situation, one would expect from the cognitive position that if there is any generalization of the effects of the VOC procedure it would be found, only for those subjects who are aware of the response-reinforcement contingency during VOC. This is inferred from their assumption that no learning takes place among subjects who are not aware (Spielberger, 1962). Surely if no learning takes place there should be no generalization. Although they have not discussed it specifically, the cognitivists to be consistent would be expected further to hold that there should be no generalization of VOC effects without subjects' awareness of a relationship between the VOC and generalization tasks. Such was not found in the present study.

The utilization in the present study of intensive interviewing procedures as suggested by Spielberger (1962) takes into account the criticism leveled against the findings of the earlier studies which found conditioning without awareness as being the result of insensitive interviewing procedures.

Actually, the writers do not interpret the present results as proving or demonstrating that "learning without awareness" has occurred. We feel that as long as the only criterion of "awareness" or "unawareness" is the retrospective report of the subject, it is impossible unequivocally to demonstrate learning without awareness, or, conversely, to prove that learning cannot occur without awareness. This is true because the subject judged aware may have his awareness either suggested by the interview itself, or it may be the result of his observing his own behavior changing as a function of the reinforcement during VOC.[2] And the critic can always claim that the subject judged unaware really was aware, but the interview was not extended enough to detect his true awareness. However, we do feel that when we utilize the criteria suggested by the cognitivists for assessing awareness, and then find that under certain conditions (e.g., with intertrial activity or other obfuscation techniques and a generalization task) our results are contrary to the cognitive predictions, that position is weakened. And our results are entirely consistent with the reinforcement interpretation of VOC.

So we would suggest that in VOC the reinforcement (if it is relevant to the motivation of the subject) strengthens the members of the reinforced response class (explainable through some mediational process of the Osgood type), and that, depending on the conditions of the experiment, the subject may or may not observe (be conscious or aware of) the change in his behavior. If he does observe the change in his own behavior, he is likely to interpret it (experience it phenomenologically) as being the result of his own conscious intentions. We have yet seen no data inconsistent with this interpretation and have seen data contradictory of the cognitive interpretation.

REFERENCES

BUSS, A. H., & DURKEE, A. Conditioning of hostile verbalizations in a situation resembling a clinical interview. *Journal of Consulting Psychology,* 1958, **6**, 415–418.

DEERING, M. F. A test of a training procedure designed to increase the intensity of angry verbalizations. Unpublished doctoral dissertation, University of Pittsburgh, 1958.

DE NIKE, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology,* 1964, **68**, 521–529.

DINOFF, M., HORNER, R. F., KURPIEWSKI, B. S., RICKARD, H. C., & TIMMONS, E. O. Conditioning verbal behavior of a psychiatric population in a group therapy-like situation. *Journal of Clinical Psychology,* 1960, **16**, 371–372.

DIXON, P. W., & OAKES, W. F. Effect of intertrial activity on the relationship between awareness and verbal operant conditioning. *Journal of Experimental Psychology,* 1965, **69**, 152–157.

DULANY, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham: Duke University Press, 1962. Pp. 102–129.

GREENSPOON, J., & THOMPSON, L. Generalization of a reinforced verbal response. Paper read at Midwestern Psychological Association, Chicago, May 1959.

INSKO, C. A. Verbal reinforcement of attitude. *Journal of Personality and Social Psychology,* 1965, **2**, 621–623.

KRASNER, L. The therapist as a social reinforcement machine. In H. H. Strupp & L. Luborsky (Eds.), *Research in psychotherapy.* Vol. 2. Washington, D. C.: American Psychological Association, 1962. Pp. 61–94.

KRUMBOLTZ, J. D., & THORESEN, C. E. The effect of behavioral counseling in group and individual settings on information-seeking behavior. *Journal of Counseling Psychology,* 1964, **11**, 324–333.

MOOS, R. H. Retention and generalization of operant conditioning effects in an interview situation. *Journal of Abnormal and Social Psychology,* 1963, **66**, 52–58.

OAKES, W. F. Verbal operant conditioning, intertrial activity, awareness, and the extended interview. *Journal of Personality and Social Psychology,* 1967, in press.

OAKES, W. F., DROGE, A. E., & AUGUST, B. Reinforcement effects on conclusions reached in group discussion. *Psychological Reports,* 1961, **9**, 27–34.

---

[2] This is how the reinforcement interpretation accounts for the commonly found relationship between degree of awareness and amount of conditioning and, indeed, how the significant interaction between awareness and conditioning found in the present study is interpreted. Naturally, the aware subjects exhibit more conditioning than the unaware, not because they must be aware to condition, but because unless they condition they have nothing of which to be aware when, in the interview, they look back at their behavior during the acquisition procedure. This is, of course, true whether the retrospective report is made after all trials or after each block of trials, as in De Nike's (1964) unsuccessful attempt to get around this difficulty inherent in the use of the restrospective report.

Rogers, J. M. Operant conditioning in a quasi-therapy setting. *Journal of Abnormal and Social Psychology*, 1960, **60**, 247–252.

Sarason, I. G., & Campbell, J. M. Anxiety and the conditioning of mildly hostile verbs. *Journal of Consulting Psychology*, 1962, **26**, 213–216.

Singer, R. D. Verbal conditioning and generalization of pro-democratic responses. *Journal of Abnormal and Social Psychology*, 1961, **63**, 43–46.

Spielberger, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness*. Durham: Duke University Press, 1962. Pp. 73–101.

Taffel, C. Anxiety and the conditioning of verbal behavior. *Journal of Abnormal and Social Psychology*, 1955, **51**, 496–501.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications, 1944.

Timmons, E. O. Experiments in conditioning operant verbal behavior. Unpublished doctoral dissertation, University of Tennessee, 1959.

Verplanck, W. S. Unaware of where's awareness: Some verbal operants—notates, monents, and notants. In C. W. Eriksen (Ed.), *Behavior and awareness*. Durham: Duke University Press, 1962. Pp. 130–158.

Weide, T. N. Conditioning and generalization of the use of affect-relevant words. Unpublished doctoral dissertation, Stanford University, 1959.

Wimsatt, R. W., & Vestre, D. N. Extraexperimental effects in verbal conditioning. *Journal of Consulting Psychology*, 1963, **27**, 400–404.

(Received August 22, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 400–407

MODELS AND HELPING:
NATURALISTIC STUDIES IN AIDING BEHAVIOR [1]

JAMES H. BRYAN AND MARY ANN TEST

Northwestern University

4 experiments concerned with helping behavior were conducted. 3 were addressed to the effects of altruistic models upon helping, while 1 was concerned with the impact of the solicitor's race upon donations. 3 investigations employed as a site parking lots of 2 large department stores in New Jersey, and indexed helping by contributions to the Salvation Army. A 4th experiment indexed helping by offers of aid by passing motorists to a woman with a disabled vehicle. Whether one employed motorists in California or shoppers in New Jersey, the results were quite consistent. The presence of a helping model significantly increased helping behavior. As race of the Salvation Army solicitor did affect the percentage of donors willing to contribute money, it was concluded that interpersonal attraction is a relevant variable affecting donations.

Recently, concern has been evidenced regarding the determinants and correlates of altruistic behavior, those acts wherein individuals share or sacrifice a presumed positive reinforcer for no apparent social or material gain. Studies addressed to these behaviors have explored both individual differences in the tendency to be altruistic and the situational determinants of such responses. Gore and Rotter (1963) found that students at a southern Negro college were more likely to volunteer for a social protest movement if they perceived sources of reinforcement as internally rather than externally guided. Subjects high on internal control were more likely to volunteer as freedom riders, marchers, or petition signers than subjects who perceived others as primary agents of reinforcement. Experimental evidence has been generated supporting the often-made assumption that guilt may serve as a stimulus to altruistic activity. Darlington and Macker (1966) found that subjects led to believe that they had harmed another through incompetent performances on the experimental tasks (three paper-and-pencil tests) were more willing than control subjects to donate blood to a local hospital. Aronfreed and Paskal [2] and Midlarsky and Bryan (1967) found that children exposed to treatment conditions designed to produce empathy were more willing to donate M&M candies than subjects given control conditions, while Handlon and Gross (1959), Ugurel-Semin (1952), Wright (1942), and Midlarsky and Bryan have found sharing to be positively correlated with age among school-age children. Lastly, Berkowitz and Friedman (1967) have demonstrated that adolescents of the working class and the bureaucratic middle class are less affected in their helping behaviors by interpersonal attraction than adolescents of the entrepreneur middle class.

Three hypotheses have emerged regarding the situational determinants of self-sacrificing behaviors. One suggests that individuals behave in an altruistic fashion because of compliance to a norm of reciprocity. That is, individuals are aware of the social debts and credits established between them, and expect that ultimately the mutual exchange of goods and services will balance (Gouldner, 1960). Berkowitz and Daniels (1964) have suggested that individuals might show a generalization of such obligatory feelings and thus aid others who had not previously assisted them.

---

[1] While Mary Ann Test collaborated with the senior author on Experiment I, the remaining work is the latter's sole responsibility.

Thanks are due to Cheryl Dellhoussay, Betty Umann, Joe McNair, and Frank Siri who served as the experimenters and stooges for Experiment I, and to Edward Nystrom, Alice Anderson, Katherine Moore, and Irene Paramoure who served as the models, observers, and solicitors in studies II, III, and IV. Studies II, III, and IV were carried out while the author was affiliated with Educational Testing Service and were supported by the National Institute of Child Health and Human Development, under Research Grant 1 PO1 HD1762-01. The authors are especially grateful to the Salvation Army of Trenton, New Jersey, and specifically to George H. Gibb, whose cooperation made these experiments possible. Thanks are also due to Perry London, David Rosenhan, Ladd Wheeler, Lawrence Stricker, and Bruce K. Eckland for the many helpful comments upon various portions of the manuscript.

[2] J. Aronfreed & V. Paskal. Altruism, empathy and the conditioning of positive affect. Unpublished manuscript, 1965.

A second hypothesis was put forth by Berkowitz and his colleagues (Berkowitz, 1966; Berkowitz & Daniels, 1963; Berkowitz, Klanderman, & Harris, 1964; Daniels & Berkowitz, 1963) who have postulated the social responsibility norm. They have contended that dependency on others evokes helping responses even under conditions where the possibility of external rewards for the helper are remote. Using supervisor's ratings of an unknown and absent other to produce dependency, and a box-construction task as the dependent variable, considerable support has been generated for the suggestion that dependency increases helping.

A third major determinant of helping may be the presence of helping (or nonhelping) models. While attention to the effects of models has generally been directed toward antisocial behaviors (cf. Bandura & Walters, 1963; Freed, Chandler, Mouton, & Blake, 1955; Lefkowitz, Blake, & Mouton, 1955), some recent evidence suggests that observation of self-sacrificing models may lead to subsequent succorant behavior by children. For example, Rosenhan and White (1967) have demonstrated that children are more likely to donate highly valued gift certificates to residents of a fictitious orphanage if they have seen an adult do so. Hartup and Coates[3] found that nursery school children who have been exposed to a self-sacrificing peer were more likely to be altruistic than children not so exposed. Test and Bryan[4] found that female college students were more likely to render aid to another in computing arithmetic problems if they saw other people so doing.

The present series of experiments was designed to test the effects of models in natural settings on subject samples other than college or high school students, and in contexts other than a school room or university setting. The first three experiments reported are concerned with the impact of observing helping models upon subsequent helping behaviors, while the fourth is addressed to the influence of interpersonal attraction upon donation behavior.

EXPERIMENT I: LADY IN DISTRESS: A FLAT TIRE STUDY

Few studies have been concerned with the effects of models upon *adults*, and fewer still with the impact of *prosocial* models upon them (Wheeler, 1966). Those that have been concerned with such behaviors have invariably employed college students as subjects. For example, Rosenbaum and Blake (1955) and Rosenbaum (1956) have found that college students exposed to a model who volunteered, upon the personal request of the experimenter, to participate in an experiment would be more likely to consent than subjects not exposed to such a model or than subjects who observed a model refuse to cooperate. Pressures toward conformity in these experiments were great, however, as the request was made directly by the experimenter and in the presence of a large number of other students.

Test and Bryan found that the observation of helping models significantly increased the subsequent offers of aid by observers. However, in that study, subjects were given the task of solving arithmetic problems and then rating their difficulty, a task ordinarily requiring autonomous efforts. Furthermore, the experiment was conducted within a university setting, a context where independence of thought is often stressed. The effects of the model may have been simply to increase the subjects' faith that assisting others was allowed. While questionnaire data of the study did not support this interpretation, such effects could not be ruled out entirely. Thus, it is possible that the model impact was simply a propriety-defining activity which reduced the inhibitions associated with such helping behavior.

In general, then, investigations of modeling that employ adults as subjects and that demand self-sacrifice on the part of subjects are limited in number, exploit strong pressures toward conformity, and rely upon college students as subjects. The present experiment was designed to assess the impact of models upon subsequent spontaneous offers of help in other than a university setting.

Method

The standard condition consisted of an undergraduate female stationed by a 1964 Ford Mustang (control car) with a flat left-rear tire. An inflated tire was leaned upon the left side of the auto. The girl, the flat tire, and the inflated tire were conspicuous to the passing traffic.

---

[3] W. W. Hartup & B. Coates. Imitation of peers as a function of reinforcement from the peer group and rewardingness of the model. Unpublished manuscript, 1966.

[4] M. A. Test & J. H. Bryan. Dependency, models and reciprocity. Unpublished manuscript, 1966.

In the model condition, a 1965 Oldsmobile was located approximately ¼ mile from the control car. The car was raised by jack under the left rear bumper, and a girl was watching a male changing the flat tire.

In the no-model condition, the model was absent; thus, only the control car was visible to the passing traffic.

The cars were located in a predominantly residential section in Los Angeles, California. They were placed in such a manner that no intersection separated the model from the control car. No turnoffs were thus available to the passing traffic. Further, opposite flows of traffic were divided by a separator such that the first U turn available to the traffic going in the opposite direction of the control car would be after exposure to the model condition.

The experiment was conducted on two successive Saturdays between the hours of 1:45 and 5:50 P.M. Each treatment condition lasted for the time required for 1000 vehicles to pass the control car. While private automobiles and trucks, motorscooters, and motorcycles were tallied as vehicles, commercial trucks, taxis, and buses were not. Vehicle count was made by a fourth member of the experiment who stood approximately 100 feet from the control car hidden from the passing motorists. On the first Saturday, the model condition was run first and lasted from 1:45 to 3:15 P.M. In order to exploit changing traffic patterns and to keep the time intervals equal across treatment conditions, the control car was moved several blocks and placed on the opposite side of the street for the no-model condition. The time of the no-model treatment was 4:00 to 5:00 P.M. On the following Saturday, counterbalancing the order and the location of treatment conditions was accomplished. That is, the no-model condition was run initially and the control car was placed in the same location that it had been placed on the previous Saturday during the model condition. The time of the no-model condition was 2:00 to 3:30 P.M. For the model condition, the control car was placed in that locale where it had been previously during the no-model condition. The time of the model condition was 4:30 to 5:30 P.M.

Individuals who had stopped to offer help were told by the young lady that she had already phoned an auto club and that help was imminent. Those who nonetheless insisted on helping her were told the nature of the experiment.

Results

The dependent variable was the number of cars that stopped and from which at least one individual offered help to the stooge by the control car. Of the 4000 passing vehicles, 93 stopped. With the model car absent, 35 vehicles stopped; with the model present, 58 halted. The difference between the conditions was statistically significant ($\chi^2 = 5.53$, corrected for continuity, $df = 1$, $p < .02$, two-tailed). Virtually all offers of aid were from men rather than women drivers.

The time of day had little impact upon the offering of aid. Fifty vehicles stopped during the early part of the afternoon; 43 during the later hours. Likewise, differences in help offers were not great between successive Saturdays, as 45 offers of aid were made on the first Saturday, 48 on the second Saturday.

The results of the present study support the hypothesis that helping behaviors can be significantly increased through the observation of others' helpfulness. However, other plausible hypotheses exist which may account for the findings. It is possible to account for the differences in treatment effects by differences in sympathy arousal. That is, in the model condition, the motorist observed a woman who had had some difficulty. Such observations may have elicited sympathy and may have served as a reminder to the driver of his own social responsibilities.

Another explanation of the findings revolves around traffic slowdown. It is possible that the imposition of the model condition served to reduce traffic speed, thus making subsequent stopping to help a less hazardous undertaking. While the time taken for 1000 autos to pass the control car was virtually identical in the model and no-model condition and thus not supportive of such an explanation, the "slowdown" hypothesis cannot be eliminated. Assuming the model effect to be real, one might still argue that it was not a norm of helping that was facilitated by the model, but rather that inhibitions against picking up helpless young ladies were reduced. That is, within the model condition, the passing motorists may have observed a tempted other and thus felt less constrained themselves regarding similar efforts. Indeed, the insistence of some people to help in spite of the imminent arrival of other aiders suggested the operation of motives other than simply helping. Indeed, while the authors did not index the frequency of pick-up attempts, it was clear that a rather large number were evidenced.

Because of the number of alternative explanations, the evidence supporting the hypothesis that the observation of helpers per se will increase subsequent aiding is weak. Experiment II was designed to test further the prediction that the perception of another's altruistic activity would elicit similar behavior on the part of the observer.

EXPERIMENT II: COINS IN THE KETTLE

The investigation was conducted on December 14th between the hours of 10:00 A.M. and 5:00 P.M. The subjects were shoppers at a large department store in Princeton, New Jersey. Observations made on the previous day indicated that the shoppers were overwhelmingly Caucasian females.

A Salvation Army kettle was placed on the sidewalk in front of the main entrance to the store. Two females, both in experimenter's employ, alternatively manned the kettle for periods of 25 minutes. One solicitor was a Negro, the other a Caucasian. Each wore a Salvation Army cape and hat. Although allowed to ring the Salvation Army bell, they were not permitted to make any verbal plea or to maintain eye contact with the passing shoppers, except to thank any contributor for his donation.

The model condition (M) was produced as follows: Once every minute on the minute, a male dressed as a white-collar worker would approach the kettle from within the store and contribute 5 cents. As the model donated, he started a stopwatch and walked from the kettle toward a parking lot as if searching for someone. He then returned to the store. The following 20-second period constituted the duration of the treatment condition.

Following a subsequent lapse of 20 seconds, the next 20-second period defined the no-model condition (NM). Within any one minute, therefore, both M and NM treatments occurred. There were 365 occasions of each treatment.

It should be noted that it was possible that some subjects in the NM condition observed the contribution of the model or a donor affected by the model. If that hypothesis is correct, however, the effects of such incidents would be to reduce rather than enhance the differences between treatments.

Results

The dependent variable was the number of people who independently donated to the Salvation Army. People obviously acquainted, as for example, man and wife, were construed as one potential donating unit. In such conditions, if both members of a couple contributed, they were counted as a single donor.

Since there were no differences in model effects for the Negro or Caucasian solicitor, data obtained from each were combined. The total number of contributors under the NM condition was 43; under the M condition, 69. Assuming that the chance distribution of donations would be equal across the two conditions, a chi-square analysis was performed. The chi-square equaled 6.01 ($p < .01$).[5]

In spite of precautions concerning the elimination of correlated observations within a treatment condition, it was possible for subjects in any one observational period to influence one another. Such influence may have been mediated through acquaintances not eliminated by our procedures or the observations of others as well as the model donating. A more conservative analysis of the data, insuring independent observation, was therefore made. Instead of comparing treatments by analyzing the number of donors, the analysis used, as the dependent variable, the number of observation periods in which there was a contribution, that is, those periods in which more than one donation occurred were scored identically to those in which only a single contribution was received. Occasions of donations equaled 60 in the M treatment, 43 in the NM condition. The chi-square equaled 2.89 ($p < .05$).

The results of Experiment II further support the hypothesis that observation of altruistic activity will increase such behavior among observers. But the matter is not yet entirely clear, for when the observer saw the model donate he saw two things: first, the actual donation, and second, the polite and potentially reinforcing interaction that occurred between the donor and solicitor. Conceivably, the observation of an altruistic model, per se, who was not socially reinforced for his behavior, would have little or no effect on an observer. The third experiment was designed to examine this possibility.

EXPERIMENT III: COINS IN THE KETTLE II

The experiment was conducted at a Trenton, New Jersey, shopping center from the hours of 10:00 A.M. to 5:00 P.M. Again, the majority of the patrons were Caucasian females. It is likely, however, that these shoppers were of a lower socioeconomic status than those in the Princeton group.

---

[5] All chi-square analyses were corrected for continuity and all tests of significance were one-tailed.

Salvation Army kettles were placed before the main entrance of a large department store (Kettle 1) and a large food center (Kettle 2). The kettles were separated by more than 200 yards. During the first 120 observations (10:00 A.M. to 12:00 P.M.), two male college students, employed by the Salvation Army and wearing its uniform, manned the kettles. The site of the experiment was Kettle 1, except on those occasions where the worker took his "coffee break." At those times, data collection was centered at Kettle 2. An equal number of M and NM conditions were run at each site, although approximately two-thirds of the observational time was spent at Kettle 1. During the remaining 240 observational periods (1:00 P.M. to 5:00 P.M.) the same male worker and his spouse alternately manned Kettle 1. The wife was stationed by the kettle for 136 minutes, the male for 104 minutes. The experiment was conducted only at Kettle 1 during the afternoon period.

Solicitors were told to make no verbal appeals for donations or responses to the model upon his contribution. While they were not informed of the hypothesis underlying the experiment, they may well have deduced it. The model was the same as in Experiment II, and again was dressed as a white-collar worker.

The imposition of the treatment conditions were identical to those described in Experiment I with the following exceptions. Since the kettle was more visible at this site than at the previous one, 30-second rather than 20-second periods were used for each treatment. To simplify the procedures, no waiting periods between treatments occurred. Additionally, after donating, the model would return to the parking lot. There were a total of 360 occasions of each of the M and NM conditions.

Results

The criteria defining a donor were identical to those outlined in Experiment I. Under the M condition, 84 donors were tallied; under the NM treatment, 56. The chi-square value was 4.86 ($p < .025$).

Since it was possible that one donor might have seen a donor other than the model receive social approval from the solicitor, the more conservative comparison of the treatments as outlined in Experiment II was made. That is, treatments were compared by noting the number of observational periods in which any donation occurred. Therefore, those donors who may have been influenced by a contributor receiving the solicitor's thanks were excluded. Of the 360 observational periods under the M condition, there were 75 in which some donation was made. Of the 360 periods, 51 were marked by contributions. Chi-square yielded a value of 5.09 ($p < .025$).

EXPERIMENT IV: ETHNOCENTRISM AND DONATION BEHAVIOR

While Experiment III was conducted to eliminate the solicitor's explicit social approval as a mechanism underlying donation behavior, it is possible that the model's impact was due to the information communicated to the observer regarding the consequence of donations. Work by Bandura, Ross, and Ross (1963), for example, found that children observing a model rewarded for aggression would be more aggressive than children who had observed a model being punished for such behavior. Additionally, considerable data have been gathered within the university laboratory suggesting that interpersonal attraction may greatly influence the helping response. Berkowitz and Friedman (1967), Daniels and Berkowitz (1963), and Goranson and Berkowitz (1966) have suggested that positive affect increases the probability of low payoff helping behavior.

The present experiment was designed to assess the impact of the solicitor's race upon the donation behavior of shoppers. It was assumed that a Negro solicitor would be held in less esteem by Caucasian shoppers than a solicitor of their same race, and that such attitudes would affect contributions. While the applicability of the "consequence to the model" hypothesis in accounting for the model's effect was not tested directly, the study assesses the importance of interpersonal attraction in eliciting charitable behavior.

Method

The experiment was conducted on December 2 and 3 between the hours of 10 A.M. and 6 P.M. at the Trenton area site. The subjects were Caucasian shoppers at a large department store.[6] Three thousand seven hundred and three shoppers were observed; 2,154 females and 1,549 males. In order to reduce the possibility of including the same subject in the experiment on more than one occasion, tallies were made only of exiting shoppers.

Two Salvation Army kettles were placed at two store exits, their location being separated by approximately 75 yards. Two female solicitors, a Negro and a Caucasian, manned the kettles. Both were in their early twenties, wore the uniform of the Salvation Army, and were in the employ of the experimenter. Each was instructed to make no verbal appeals for donations and to avoid eye contact with the shoppers. After a period of 25 minutes, the girls rotated kettle assignments, and during the last 10 minutes of the hour were allowed to take a coffee break. Hence, during a single hour, each solicitor manned both kettles. Each solicitor manned each kettle on seven occasions per day. Thus, each solicitor was observed for a total of 28 observational periods; 14 on each day (seven on each kettle) over a period of two days.

---

[6] As there were very few Negro donors ($N = 7$), analysis was confined to the behavior of Caucasian shoppers.

Two observers, each assigned to a particular kettle, tallied the number and sex of the exiting shoppers and contributors during each of the 25-minute periods. In addition, records were kept of the amount of money donated within any period, although it was impossible on this measure to separate those donations made by incoming from outgoing customers.

Results

The dependent variable was the percentage of donors contributing to the kettle within an observational period. That is, observational periods were assigned a percentage donor score. Shoppers within an observational period were treated as a single group, with differences between groups on percentage donor score forming the critical comparisons. The total *N* of the study was then the 56 observational periods, rather than the 3,703 shoppers. Since the mean group size for the Negro solicitor was 70.32 and for the Caucasian 61.93 (standard deviations equal to 53.33 and 42.98, respectively), it was assumed that the percentage score was relatively stable.

The effects of race, kettle location, and day and their interactions were analyzed by analysis of variance.

As can be seen from Table 1, both the main effect of race and of day were significant. As predicted, the Negro solicitor elicited a statistically significant lower percentage of donors than did the Caucasian. For the Negro solicitor, the average percentage donor score for observational periods was 2.22 ($SD = 2.36$), while for the Caucasian solicitor the average percentage donor score was 3.89 ($SD = 3.60$). Additionally, Saturday shoppers were by and large less generous than Friday customers. The average percentage donor score of the group was 1.73 ($SD = 1.97$) for the Saturday shopper, and 4.38 for the Friday shopper ($SD = 3.52$).

A second dependent variable was the amount of money donated during each time period. No significant differences were found for race, day, or kettle location.

The present investigation does support, albeit equivocally, the notion that interpersonal attraction may affect donations even when the solicitors are not the eventual recipients of such contributions. While it is possible that race differences simply fail to remind observers of their social responsibilities, it is also feasible that the subjects wanted to avoid interpersonal contact with a minority group member. If this is true, then it is interesting to note that interpersonal attraction may play an important role even in those situations where personal anonymity is high and escape from unpleasant situations easy.

TABLE 1

ANALYSIS OF VARIANCE OF PERCENTAGE DONOR SCORES

	df	*MS*	*F*
Race (A)	1	38.778	4.84*
Day (B)	1	98.315	12.28**
Kettle (C)	1	.018	
A × B	1	1.511	
A × C	1	11.340	
B × C	1	1.031	
A × B × C	1	3.206	
Error	48	8.009	

* $p < .05$ (2-tailed).
** $p < .01$ (2-tailed).

DISCUSSION

The results of the first three experiments clearly replicate those of Test and Bryan and extend the findings over a variety of subject populations, settings, and tasks. The results hold for college students, motorists, and shoppers; in the university laboratory, city streets, and shopping centers; and when helping is indexed by aiding others solve arithmetic problems, changing flat tires, or donating money to the Salvation Army. The findings then are quite consistent: the presence of helping models significantly increases subsequent altruistic behavior.

That generosity breeds generosity is interesting in light of the recent concern with helping behaviors in emergency contexts. Darley and Latané[7] and Latané and Darley[8] have found that subjects are less inclined to act quickly in emergency situations when in the presence of other potential helpers. Whether faced with a medical emergency (a simulated epileptic seizure) or a dangerous natural event (simulated fire), the rapidity with which students sought to aid was reduced by the presence of others. These findings have been interpreted in three ways: as reflecting the subjects' willingness to diffuse responsibility (others will aid); as reflecting their diffusion of blame (others didn't aid either); or as reflecting conformity to the nonpanicked stooges. It is clear that the results of the first three experiments in the present series do not follow that which might be predicted by the diffusion concepts. A giving model apparently does not lend credibility to the belief that others than the self will make the necessary sacrifices. The helping other did not strengthen the observer's willingness to diffuse his social obligations, but rather stimulated greater social responsibility. In light of these results, the delayed reaction exhibited by the subjects tested by Darley and Latané might be best attributable to conformity behavior. As they have suggested, subjects faced with a unique and stressful situation may have been either reassured by the presence of calm others or fearful of acting stupidly or cowardly. Additionally, it is possible that diffusion of responsibility is only associated with anxiety-inducing situations. The current data fail to indicate that such diffusion occurs in nonstressful situations which demand fulfillment of social obligations.

---

[7] J. Darley & B. Latané. Diffusion of responsibility in emergency situations. Unpublished manuscript, 1966.

[8] B. Latané & J. Darley. Group inhibition of bystander intervention in emergencies. Unpublished manuscript, 1966.

While it appears clear that the behavior of the motorists and shoppers was not dictated by a variety of situational and social pressures usually associated with the study of modeling in adults or experiment in academic settings (Orne, 1962), the mechanisms underlying the effects are not obvious. While the presence of the model in the flat-tire study may have reminded the motorists as to the social responsibility norm, a hypothesis does not appear reasonable in accounting for the results in the coins-in-the-kettle series. The bell-ringing Salvation Army worker, with kettle and self placed squarely in the pathway of the oncoming pedestrian, would seem to be reminder enough of one's obligation toward charity. A priori, it would not appear necessary to superimpose upon that scene the donating other for purposes of cognitive cueing (Wheeler, 1966).

One hypothesis to account for the model effect is that the observer is given more information regarding the consequences of such donation behavior. Experiment IV suggested that solicitor status or personal attraction might operate on donation behaviors even under conditions of personal anonymity and few social constraints. It is possible that the model serves to communicate to the potential donor relevant information concerning the consequences of his act. That is, the model may demonstrate that an approach to the solicitor does not involve an unwanted interpersonal interaction (e.g., lectures on religion).

A second hypothesis to account for the data pertains to the shame-provoking capacities of the model. It is reasonable to assume that most people feel that they are, by and large, benevolent and charitable. Furthermore, it is likely that such a self-image is rarely challenged: first because charitable acts are not frequently required; second, at least in the street scenes employed in the current series of studies, solicitations are made in the context of many nongiving others. That is, a multitude of negative models—of noncharitable others—surround the solicitations in the current series of studies. Indeed, the contexts are such that most people are not helping; many more cars pass than stop to offer aid to the lady in distress; and there are many more people who refuse to put coins in the kettle than those who do. However, the witnessing of a donor, an individual who not only recognizes his social responsibility but in fact acts upon it, may produce a greater challenge to the good-self image of the observer. Acts rather than thoughts may be required of the observer in order to maintain the self-image of benevolence and charity. If such is the case, then the model characteristics most effective in producing prosocial behavior by socialized adults would be those directed toward shame or guilt production (e.g., donations from the poor), rather than those reflecting potential reinforcement power (e.g., donations from the high status).

Whatever the mechanism underlying the model effect, it does appear quite clear that prosocial behavior can be elicited through the observation of benign others.

REFERENCES

BANDURA, A., ROSS, D., & ROSS, S. Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*, 1963, **66**, 601–607.

BANDURA, A., & WALTERS, R. H. *Social learning and personality development*. New York: Holt, Rinehart & Winston, 1963.

BERKOWITZ, L. A laboratory investigation of social class and national differences in helping behavior. *International Journal of Psychology*, 1966, **1**, 231–240.

BERKOWITZ, L., & DANIELS, L. Responsibility and dependency. *Journal of Abnormal and Social Psychology*, 1963, **66**, 429–436.

BERKOWITZ, L., & DANIELS, L. Affecting the salience of the social responsibility norm: Effects of past help on the response to dependency relationships. *Journal of Abnormal and Social Psychology*, 1964, **68**, 275–281.

BERKOWITZ, L., & FRIEDMAN, P. Some social class differences in helping behavior. *Journal of Personality and Social Psychology*, 1967, **5**, 217–225.

BERKOWITZ, L., KLANDERMAN, S. B., & HARRIS, R. Effects of experimenter awareness and sex of subject and experimenter on reactions to dependency relationships. *Sociometry*, 1964, **27**, 327–337.

DANIELS, L., & BERKOWITZ, L. Liking and response to dependency relationships. *Human Relations*, 1963, **16**, 141–148.

DARLINGTON, R. B., & MACKER, C. E. Displacement of guilt-produced altruistic behavior. *Journal of Personality and Social Psychology*, 1966, **4**, 442–443.

FREED, A., CHANDLER, P., MOUTON, J., & BLAKE, R. Stimulus and background factors in sign violation. *Journal of Personality*, 1955, **23**, 499.

GORANSON, R., & BERKOWITZ, L. Reciprocity and responsibility reactions to prior help. *Journal of Personality and Social Psychology*, 1966, **3**, 227–232.

GORE, P. M., & ROTTER, J. B. A personality correlate of social action. *Journal of Personality*, 1963, **31**, 58–64.

GOULDNER, A. The norm of reciprocity: A preliminary statement. *American Sociological Review*, 1960, **25**, 161–178.

HANDLON, B. J., & GROSS, P. The development of sharing behavior. *Journal of Abnormal and Social Psychology*, 1959, **59**, 425–428.

LEFKOWITZ, M., BLAKE, R., & MOUTON, J. Status factors in pedestrian violation of traffic signals. *Journal of Abnormal and Social Psychology*, 1955, **51**, 704–706.

MIDLARSKY, E., & BRYAN, J. H. Training charity in children. *Journal of Personality and Social Psychology*, 1967, **5**, 408–415.

ORNE, M. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776–783.

ROSENBAUM, M. The effect of stimulus and background factors on the volunteering response. *Journal of Abnormal and Social Psychology*, 1956, **53**, 118–121.

ROSENBAUM, M., & BLAKE, R. Volunteering as a function of field structure. *Journal of Abnormal and Social Psychology*, 1955, **50**, 193–196.

ROSENHAN, D., & WHITE, G. M. Observation and rehearsal as determinants of prosocial behavior. *Journal of Personality and Social Psychology*, 1967, **5**, 424–431.

UGUREL-SEMIN, R. Moral behavior and moral judgment of children. *Journal of Abnormal and Social Psychology*, 1952, **47**, 463–474.

WHEELER, L. Toward a theory of behavioral contagion. *Psychological Review*, 1966, **73**, 179–192.

WRIGHT, B. A. Altruism in children and perceived conduct of others. *Journal of Abnormal and Social Psychology*, 1942, **37**, 218–233.

(Received September 15, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 408–412

INDIVIDUAL DIFFERENCES AND PERCEPTUAL DEFENSE IN THE ABSENCE OF RESPONSE BIAS[1]

RICHARD R. BOOTZIN AND MARK W. STEPHENS

Purdue University

40 college students participated in a tachistoscopic perceptual recognition task. A forced-choice procedure was used which minimized the effect of response biases upon perceptual accuracy. Affective connotations of the stimuli (random forms) were established through success-failure manipulations. All Ss were given a personality-test battery in order to identify "repressors" and "sensitizers." Although the specific hypotheses of the experiment were not confirmed, additional support was obtained for the more general hypothesis that the affect of a stimulus influences its perceptibility. A perceptual sensitization effect was manifest for the 1st $\frac{1}{3}$ of the experiment. The personality scales failed, however, to predict individual differences with respect to this effect.

Perceptual defense research has undergone substantial methodological improvement, but still there is no consensus as to whether the perceptual defense phenomenon in fact exists and, if so, to what variables it is related. The present study was intended as an effort to reexamine the phenomenon, making use of the methodological sophistication gained in recent years.

There have been methodological improvements in all three of the critical elements in a perceptual defense study: the stimuli, the responses, and the subjects. The majority of earlier studies used words as stimuli—either "dirty" words, or words selected by a word-association test as having apparent emotional connotation to the subject. Words may not be sufficiently anxiety arousing to elicit perceptual defense, and, as Solomon and Howes (1951) suggested, dirty words may simply be unfamiliar visual stimuli. Poorer recognition of them may simply be a function of lesser familiarity. Nonverbal stimuli have been used less frequently, and when they have been used there are often questions regarding equivalence of recognizability.

Questions regarding the *response* in perceptual defense research have centered on the problem of response bias or response suppression. Early studies relied primarily on the ascending method of limits to determine thresholds. This method can produce artifactual results any time the subject is reluctant to give a response. For example, he may, simply, be more inclined to "wait until he is sure" before giving a taboo response. Goldstein (1962, 1964) and others have offered evidence that all "perceptual defense" phenomena reported using this method may be artifactual.

Concern with *subject* variables dates from the original Bruner and Postman (1947) study in which some subjects showed perceptual defense and others perceptual vigilance. If data from different subjects are pooled, the result may be no apparent evidence of either perceptual defense or perceptual vigilance. Several people, most recently Byrne (1961), have developed tests designed to measure the tendency to use perceptual defense (or repression) and perceptual vigilance (or sensitization). There has been some success in finding these tests related to differences in perceptual defense as manifested in controlled laboratory perceptual tasks. However, the studies using these tests have often permitted room for artifacts attributable to problems involved with the stimulus or the response.

In the present study response bias was controlled by a forced-choice stimulus-presentation technique, differential familiarity or recognizability was controlled by using random forms of equal recognizability, and individual differences in use of perceptual defense were investigated by means of Byrne's (and others') tests designed to measure this variable. It was expected that "sensitizers" (as inferred from Byrne's Repression-Sensitization Scale) would manifest perceptual vigilance and "repressors" would manifest perceptual defense.

---

[1] This study met part of the requirements for a master's degree (Bootzin, 1966) and was supported by Predoctoral Research Fellowship FI MH-24, 655-02 of the Public Health Service of the National Institutes of Health awarded to the senior author.

METHOD

Subjects

The subjects were 40 male freshmen enrolled in a freshman-sophomore general psychology course, fulfilling a course requirement by participating in the experiment.

Perceptual Task

The stimuli were eight-sided random forms of the type used by Attneave and Arnoult (1956) and others (see Figure 1). A pilot study (Bootzin, 1966) determined that the forms were of equivalent recognizability.

A forced-choice response procedure was used to control response-bias effects. On each trial, a photograph of one of the forms was projected tachistoscopically. As soon as the stimulus was flashed, the subject indicated which of two forms on his answer sheet had been projected. (As a further test for response biases, pseudo-accuracy scores were calculated for each subject, using four-sided random forms projected at .01 second as stimuli, but the same eight-sided forms as response alternatives. Using a randomly selected set of responses as "correct responses," an analysis of variance was performed on pseudo-accuracy scores. There were no significant main effects or interactions in pseudo-accuracy scores.) To compare the treatment effects that occur above threshold, as well as those that occur near threshold, the forms were presented at two exposure speeds, .01 and .04 second. Each subject was given six trials per form per exposure time, or 96 trials in all.

Induction of Negative Affect

Negative affect was associated with half the stimuli by means of a preexperimental task in which the stimuli became associated with failure. At the beginning of the session, all subjects completed a fabricated intelligence test called the Stanford Form Test, consisting of 10 puzzles. Each of the eight-sided forms used later in the recognition task had been divided into five or six parts and each part had been reproduced onto transparent acetate. The task consisted of placing the transparencies on top of one another so that they reconstructed the form, with a 1-minute time limit. To equalize time sent scanning the forms, if a subject finished in less than 1 minute, he recorded his time with the clock stopped, then restarted the clock, reshuffled the transparencies, and began the puzzle again. The puzzles were constructed so that 6 were solvable and 4 were unsolvable. To avoid the possibility that a difference in perceptual accuracy between "failure" and "success" stimuli was due to different populations of stimuli, each stimulus was a failure stimulus for half the subjects and a success stimulus for the other half.[2]

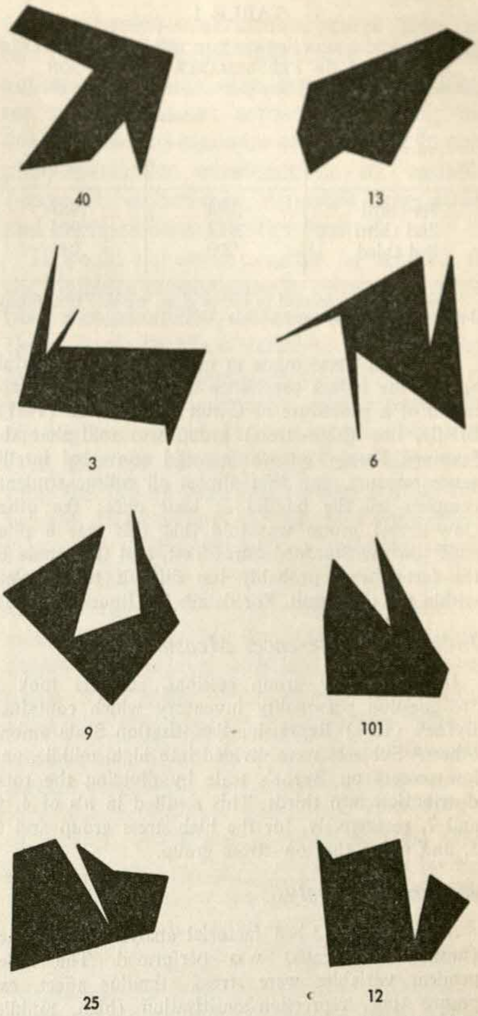


FIG. 1. The random forms used in the Stanford Form Test and in the perceptual recognition task.

---

[2] In a pilot study with these puzzles, subjects sometimes did not complete the solvable ones within the time limit. Therefore, a line was added to the descriptive outlines of the five most difficult forms which made them easier to solve by indicating where one piece would fit. Similar lines were added to some of the outlines for unsolvable forms so that the subjects would not associate success with the presence of a partition. Still, 10 of 40 subjects failed at least one of the puzzles that were solvable, and some puzzles were failed more often than others. Even so, the stimulus populations were substantially the same (Bootzin, 1966).

TABLE 1

MEAN RECOGNITION SCORES FOR THE STIMULUS AFFECT × STAGE OF PERFORMANCE INTERACTION

Stage of performance	Stimulus affect	
	Success	Failure
1st third	.884	.987
2nd third	.987	.962
3rd third	.990	.941

Note.—Units expressed in terms of an arc-sine transformation; chance performance equals .785.

An attempt was made to manipulate the stressfulness of the failure experience by means of an adaptation of a procedure of Caron and Wallach (1957). Briefly, one (high-stress) group was told that the Stanford Form Test was a valid nonverbal intelligence measure, and that almost all college students complete all the puzzles at least once; the other (low-stress) group was told that this was a pilot study on the Stanford Form Test, and that some of the forms were probably too difficult to complete within the time limit. For details see Bootzin (1966).

Individual Differences Measures

In preliminary group sessions, subjects took a 385-question personality inventory which contained Byrne's (1961) Repression-Sensitization Scale among others.[3] Subjects were divided into high, middle, and low scorers on Byrne's scale by dividing the total distribution into thirds. This resulted in *n*'s of 4, 9, and 7, respectively, for the high-stress group and 9, 5, and 6 for the low-stress group.

Design and Analysis

A 2 × 2 × 2 × 3 × 3 factorial analysis of variance (unweighted means) was performed. The independent variables were stress, stimulus affect, exposure time, repression-sensitization (high, middle, and low thirds of the distribution on Byrne's scale), and stage of performance (first, second, and third third of the trials). Stimulus affect, exposure time, and stage of performance were repeated-measures factors.

The relation between the above individual difference measures and perceptual accuracy was obtained through correlations and regression analyses. Because of space limitations these analyses are not included here, but can be found in Bootzin (1966).

The dependent variable was the proportion of correct responses per subject; the raw proportion scores were converted to angle scores by an arc-sine transformation. With this transformation, a chance-level accuracy score (.50) was equal to a transformed score of .785.

RESULTS

The general question of this study was whether the prior association of a stimulus with a success or failure experience influences perceptual accuracy. Accuracy was significantly greater on failure than on success stimuli during the first third of the trials (Stage of Performance × Stimulus Affect interaction, $F = 7.085$, $df = 2, 68$, $p < .01$); this effect did not persist, however (stimulus affect main effect, $F < 1.0$). The means are presented in Table 1. It was expected that performance would improve across trials. Examination of Table 1, however, indicates that performance did improve on success stimuli, but not failure stimuli. In brief, there was perceptual sensitization during early trials, which dissipated in later trials.

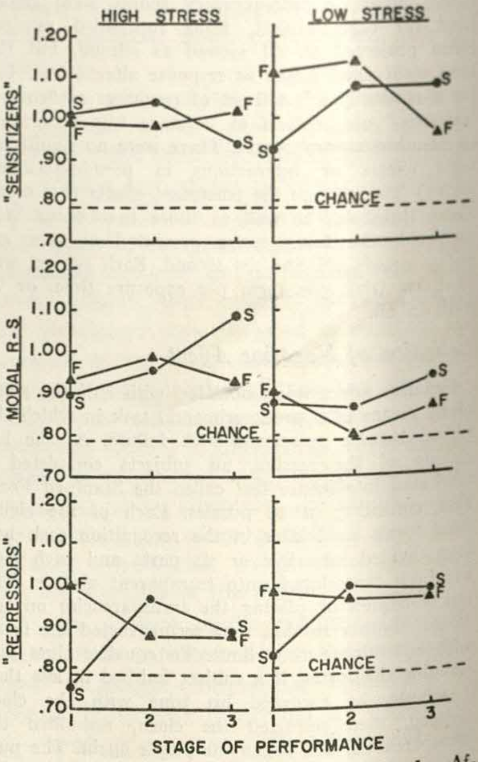


FIG. 2. The Repression-Sensitization × Stimulus Affect × Stress × Stage of Performance interaction.

---

[3] The 11 other scales that were administered were the Hysteria (*Hy*), Psychasthenia (*Pt*), and Suppressor (*K*) scales of the MMPI, the Taylor Manifest Anxiety scale, Welsh's *A* and *R* scales, Edwards' Social Desirability scale, the Marlowe-Crowne Social Desirability scale, a Locus of Control scale (Rotter, 1966), and Edwards' Aggression and Endurance scales from Edwards' Personal Preference Schedule. These last two scales defined the opposite poles of the factor preservation for achievement in Caron and Wallach (1959).

Contrary to expectation, there was no tendency for sensitizers, as determined from Byrne's scale, to perform better on failure items and/or less well on success items than repressors (Repression-Sensitization × Stimulus Affect, $F < 1.0$). None of the interactions involving repression-sensitization and stimulus affect reflected the expected result. The only such interaction that approached significance was Repression-Sensitization × Stimulus Affect × Stress × Stage of Performance ($F = 2.285$, $df = 4, 68$, $p < .10$). Even this highly complicated interaction (see Figure 2) was, if anything, opposite to expectation. The repressors were perceptually vigilant in both high- ($p < .01$) and low-stress ($p = .07$) conditions, while the sensitizers were vigilant only in the low-stress ($p < .10$) condition.

In general, however, sensitizers performed better than repressors on all stimuli, regardless of stimulus affect (repression-sensitization, $F = 4.097$, $df = 2, 34$, $p < .05$). This generalization is also restricted, of course, by a significant interaction involving repression-sensitization (Repression-Sensitization × Stress × Exposure Time × Stage of Performance, $F = 4.000$, $df = 4, 68$, $p < .01$).

Results concerning exposure time also reflected complex effects of specific combinations of variables although, generally, performance was, as expected, better at .04 second than at .01 second (exposure time, $F = 30.742$, $df = 1, 34$, $p < .001$). There were three significant interactions involving exposure time: Exposure Time × Stage of Performance ($p < .05$), Stress × Exposure Time × Stage of Performance ($p < .05$), and Repression-Sensitization × Stress × Exposure Time × Stage of Performance ($p < .01$). In all of these, recognition was consistently better at the longer exposure time, but the magnitude of difference between exposure times was a function of a combination of other variables.

DISCUSSION

The important findings can be summarized as follows: (*a*) There was clear evidence of perceptual vigilance uncontaminated by response bias, although the vigilance was apparent only for the first third of the trials; (*b*) repression-sensitization scores were not related to individual differences in defense/vigilance, although sensitizers performed better than repressors across all stimuli; and (*c*) defense and vigilance were related to complex, particular combinations of variables (stage of performance, exposure time, stress, and repression-sensitization scores).

It would not seem possible to account for the vigilance phenomenon in terms of response bias, recognizability, or familiarity because of the controls on these variables in the design, and analysis of pseudo-accuracy scores revealed no pseudo-accuracy effects paralleling the vigilance effect. It remains possible, however, that the subjects spent more time scanning the failure than the success stimuli during the affect-induction task and were therefore more familiar with them as visual stimuli during the recognition task. (It was to control familiarity that subjects were instructed to reshuffle and reassemble success forms after first assembling them, to assure equal time spent with success and failure stimuli.) However, the correlation between recognition performance and time of first successful completion of success puzzles was not significantly different from zero ($r = .095$, $p > .25$). In other words, there was no tendency for the hard success puzzles to be better recognized than the easy success puzzles. Consequently, it would not seem likely that the better recognition of failure stimuli can be adequately accounted for as a result of greater prior scanning time and, thus, familiarity.

The finding that the vigilance effect persisted only through the first third of the trials suggests that perceptual defense and vigilance may be fairly fragile phenomena which may dissipate quickly as a function of new experience with the stimuli. Bootzin and Natsoulas (1965), using somewhat different procedures, found a perceptual defense phenomenon which, similarly, persisted for only the first half of the trials. These phenomena may be less fragile when the subject has more extended and/or more intense experience with the affect-associated stimuli than that provided by this experimental paradigm. However, it is also possible that defense or vigilance responses to particular stimuli are fairly situationally specific and may extinguish rapidly when contingencies are changed.

The failure of repression-sensitization scores, or scores on 11 other personality tests, to be related to individual differences in defense or vigilance behavior must be interpreted with caution because of the small number of subjects per cell, even though the primary test of personality correlates was based on the complete sample of 40 subjects. However, the finding of particularistic effects of combinations of variables (including personality variables) suggests a different conception of correlates of individual differences in use of perceptual defense. Perceptual defense and vigilance may be relatively specific responses, learned to specific classes of stimuli and situations, rather than aspects of more general personality characteristics; subjects may not differ so much in their generalized tendency to repress or sensitize as in the specific classes of stimuli and/or situations in which they have learned to employ defense or vigilance. Thus, in some situations most individuals may perceptually defend, while in others most may show perceptual vigilance. Phares (1962) found results consistent with this interpretation. In the present study, the affect induced was failure in an academic-type task, in which selective attention to failure may well be functional—and thus ordinarily produce vigilance. In other studies, affect is sometimes inferred from emotional responses to a word-association test. Perceptual defense, in the latter studies, may be more functional. Bootzin and Natsoulas (1965) using a word-association test did, in fact, find that most of their subjects were defenders.

The major conclusion of the present study is that even when stimulus, response, and subject variables are well controlled, a perceptual defense (or vigilance) phenomenon can be found. However, the situational cues and subject characteristics which determine whether defense, vigilance, both, or neither will be elicited still have tc be specified.

REFERENCES

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychological Bulletin,* 1956, **53**, 452–471.

BOOTZIN, R. Individual differences and perceptual defense in the absence of response bias. Unpublished master's thesis, Purdue University, 1966.

BOOTZIN, R. R., & NATSOULAS, T. Evidence for perceptual defense uncontaminated by response bias. *Journal of Personality and Social Psychology,* 1965, **1**, 461–468.

BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *Journal of Personality,* 1947, **16**, 69–77.

BYRNE, D. The Repression-Sensitization Scale: Rationale, reliability, and validity. *Journal of Personality,* 1961, **29**, 334–349.

CARON, A. J., & WALLACH, M. A. Recall of interrupted tasks under stress: A phenomenon of memory or of learning? *Journal of Abnormal and Social Psychology,* 1957, **55**, 372–381.

CARON, A. J., & WALLACH, M. A. Personality determinants of repressive and obsessive reactions to failure-stress. *Journal of Abnormal and Social Psychology,* 1959, **59**, 236–245.

GOLDSTEIN, M. J. A test of the response probability theory of perceptual defense. *Journal of Experimental Psychology,* 1962, **63**, 23–28.

GOLDSTEIN, M. J. Perceptual reactions to threat under varying conditions of measurement. *Journal of Abnormal and Social Psychology,* 1964, **69**, 563–567.

PHARES, J. E. Perceptual threshold decrements as a function of skill and chance expectancies. *Journal of Psychology,* 1962, **63**, 399–407.

ROTTER, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs,* 1966, **80**(1, Whole No. 609).

SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychological Review,* 1951, **58**, 256–270.

(Received September 5, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 413-428

BEHAVIORAL RECEPTIVITY TO DISSONANT INFORMATION [1]

TIMOTHY C. BROCK AND JOE L. BALLOUN
Ohio State University *University of California, Berkeley*

The dependent variable was the amount of button pressing to produce momentary elimination of the static that partially masked tape-recorded messages. In 4 experiments, employing 112 undergraduates, it was repeatedly found that: smokers pressed more than nonsmokers to remove static from a message disputing the link between smoking and lung cancer; smokers made fewer attempts than nonsmokers to clarify a message affirming a smoking-cancer link; persons having considerable recourse to prayer and church attendance pressed less to clarify a message attacking Christianity than persons weakly committed to religion. Novelty, utility, relevance, and other factors affecting information receptivity were empirically analysed. It was concluded that dissonance theory adequately handles selective *attention* but that prediction of selective *exposure*, receptivity to future nonsupportive messages, requires taking into account the amount of supportive information already assimilated.

More adequate explanation and prediction of the effectiveness of communicative acts will be possible if selective self-exposure by prospective recipients can be better understood. Differential receptivity to messages, an ubiquitous aspect of communication processes, was a problem entirely bypassed by Hovland and the Yale group: ". . . all of the studies are concerned with problems encountered when an audience is available to a communicator and do not deal with the *prior problem* of securing . . . the audience [Hovland, Janis, & Kelley, 1953, p. 12]." More recently it has been recognized that experimentation dealing directly with self-exposure would help bridge the gap (Hovland, 1959) between laboratory and survey studies of communication and persuasion. The resulting efforts produced a literature, adequately reviewed elsewhere (Brock, 1965; Freedman & Sears, 1965), which dealt mainly with the hypothesis that persons will seek out supportive information and avoid nonsupportive information. When factors other than sheer supportiveness of the information were analytically or experimentally controlled, evidence for the widely assumed (c.f. references in Brock, 1965) avoidance of nonsupportive information has been almost entirely lacking (e.g., Freedman, 1965; Mills, 1965a, 1965b; Sears, 1965). The recurrence of negative or ambiguous findings led Freedman and Sears (1965) to suggest that other investigators "turn away from questions dealing primarily with the selective exposure hypothesis [p. 94]." Their despairing but well-documented advice might prove to be sagacious strategy if, indeed, over the long term, nothing further can be done to provide theoretical and methodological clarification concerning information supportiveness as a determinant of receptivity. However, the present report, which describes four experiments employing a novel behavioral measure of information receptivity, presents data that are apparently congenial with the theoretically postulated (Festinger, 1957) "avoidance of dissonant information" and not easily explainable by rival formulations.

Selective exposure may have become a troubled area in social psychology research in part because methodology for measurement of the dependent variable has been insensitive to the actual tuning behavior of the recipient. In most previous studies the individual indicated his preferences among messages (e.g., Brock, 1965; Freedman, 1965; Mills, Aronson, & Robinson, 1959); in a few studies (Jecker, 1964; Sears & Freedman, 1965) the experimenter managed to record how much time the subject spent reading various messages. It is unclear why preference ratings of supportive and nonsupportive information, and to a lesser extent, exposure time, have been so invariably preferred as dependent measures. Surely there was *no theoretical necessity:* perhaps later investigators mindlessly adopted the dependent measures selected by the pioneers (Ehrlich, Guttman, Schonbach, & Mills, 1957; Festinger, 1957, pp. 162–176; Mills et al., 1959) and perhaps the discouraging ambiguity in the literature (Freedman & Sears, 1965) may be partially attributable to these insufficiently sensitive measures.

---

[1] Supported by GS-343 and GS-606, grants to the first writer from the National Science Foundation. The data were collected while the writers were at Iowa State University. The writers are greatly indebted to David O. Sears for his helpful criticisms. The comments of Lee Alan Becker, Anthony G. Greenwald, and Jon D. Jecker, are gratefully acknowledged.

In real life, prospective recipients *tune out* messages which might impinge by inattention, by allowing other messages to supervene, by avoidance of the message source, by attempting to escape the context in which the message would surely be received, or by attempting to change the source so that it emits more desirable messages. In real life, prospective recipients may *tune in* impinging messages by heightening attention, by eliminating inputs which would mask or conceal the message, by seeking out the message source, by maintaining a role in the context in which the message would surely be received, or by encouraging the source so that it continues to emit acceptable messages. The foregoing taxonomy of information receptivity is far from complete; it is, however, suggestive of the range of behaviors which might plausibly serve as dependent variables in experimental study of the determinants of information receptivity.

Before deciding that this or that hypothesis about selective exposure should be abandoned, further experimental work can be directed at the many unexplored aspects of the phenomenology of tuning behavior. The important admonishments of Webb, Campbell, Schwartz, and Sechrest (1966) concerning "truth" and multiple methods should be heeded. The present measure of information receptivity was prompted by observation of persons who are eager to eliminate interfering static from a highly desired audio presentation, as, for example, a representative American male tuning his radio for a heavyweight boxing championship broadcast. In his discussion of a behavioral measure of television viewing, Lindsley (1962) summarized the Skinnerian operant literature, particularly studies of conjugate reinforcement (the intensity of a continuously available reinforcing stimulus varies directly and immediately with the rate of response); he described a switch-press response somewhat comparable to the static-eliminating button-press response used here. The present hypothesis was that persons will press more frequently to eliminate masking static from consonant than from dissonant messages. The hypothesis was tested by exposing persons to a series of auditory messages accompanied by static and measuring the frequency of static-eliminating button pressing elicited by the messages. By adjusting the static so that it was masking but never noxious, differences in the consonance-dissonance of the messages might be reflected in static removal behavior. Six communications were designed to permit comparisons among messages that varied in dissonance, interest, relevance, and intelligibility. Each of four experiments employed the same messages and dependent variable; the last three of these experiments facilitated assessment of the generality and theoretical clarity of the relationships by varying the listeners' set, the message volume, and the order of presentation of the messages.

EXPERIMENT I

Method

Messages. Pilot work and previous experience (Brock, 1962, 1965) led to the decision to manipulate dissonance naturalistically by taking account of the subjects' chronic commitments to smoking and religious behavior and then exposing the subjects to messages which varied in supportiveness for smoking, prayer, and church attendance.[2] The messages, each 5 minutes in length, were presented in the following order. *Educational Films,* a dull recitation of the titles and playing time of some educational films, was taken from a film catalogue; pilot work indicated the need for a "filler" message while subjects adjusted to the listening situation. *Cancer Link,* an authoritative presentation of the evidence linking smoking to death from lung cancer, was taken from a booklet published by the American Cancer Society (1963). *Indian Neutrality,* a plea for Indian neutrality, was spoken in Tamil,[3] a language incomprehensible to every subject. *Christianity is Evil,* an attack on the hypocrisy and wrongdoing of organized Christian religion, was based on atheistic propaganda by Berkman (1920, pp. 72–74). *No Cancer Link,* an authoritative refutation of the arguments linking smoking to lung cancer, was taken from Northrup (1957) and Sanford (1964). *No Draft for College Graduates,* taken from Bures (1963), advocated that college graduates should be draft-exempt. As part of the procedure in Experiment I, the subject was given a list of the topics to read before exposure to the speeches. The list included an appropriate title for the speech, a statement of the principal conclusion, and identification of the speaker's nationality.[4] Thus, the subject knew the nature of the speeches in advance, particularly that he would hear messages both opposing and supporting the smoking-cancer link.

---

[2] Another strategy would require the subject to make a choice and then react to messages supporting the chosen and rejected alternatives (e.g., Jecker, 1964; Mills, 1965b). Postdecisional dissonance should produce differential receptivity to supportive and nonsupportive communications. Work employing behavioral measures of receptivity in the postdecisional paradigm is in progress at the first writer's laboratory.

Instructions to subjects. The listeners were given the following printed instructions.

> College students are asked to judge the persuasiveness and sincerity of talks prepared by high school juniors and seniors. The talks you will hear were prepared for oral presentation by seniors and juniors at Iowa high schools. They will be spoken by a male American speaker or by a male foreign speaker. Adjust the earphones so they fit and listen to the series of brief persuasive talks. After each talk you will be given a rating sheet to rate the persuasiveness and sincerity of the speech. Since the talks were recorded on a small portable tape recorder there is considerable electrical interference. The interference can be "adjusted out" by pressing and then immediately releasing the control button. Use of the control several times in a row reduces somewhat the static and other interference noise.

After each of the six talks, the subjects rated the persuasiveness and sincerity of the speakers as described below. Subjects' ratings of the speeches were ostensibly the data of primary interest to the experimenter.

Apparatus and procedure. There were two concomitant auditory inputs to the subject: static and spoken message. A tape recording of static was made by connecting directly into a radio tuned to a signal-free frequency at about 800 kilocycles. Both recorder and radio were about 6 feet from a 15-watt fluorescent light; considerable natural static was recorded in combination with that generated by the light. The static was played to the listener on a Model 464 Sony tape recorder with volume set at 6.5 for all four experiments. Decibel values, based upon characteristics of the electrical output to the subjects' earphones, ranged between 33.1 and 34.1 during the 5-minute static play. The spoken messages were played to the listener on a Model T-2000 Revere tape recorder with volume set at 4.5 for the first three experiments, and 3.5 for the fourth experiment. Decibel values for the speeches ranged between 32.5 and 34.1 with the average range for a particular speech about 0.7. Reducing the volume setting for the spoken messages from 4.5 to 3.5 in the fourth experiment lowered the decibel range to 30.7–32.7. This reduction in Experiment 4 will be explained. By pressing a button five times, 3 seconds of static-free communication were obtained, that is, the static input was interrupted for 3 seconds.

Measures and treatment of data. Button presses were recorded as pen deflections on a moving chart. The basic datum was the number of deflections occurring in the first complete 15 seconds of each of the 5 minutes required for a communication. The basic datum was subjected to a square-root transformation (Mosteller & Bush, 1954, pp. 326–327) because there was marked correlation between means and variances.

In order to maintain the guise of the research, subjects rated each communication immediately afterwards. A 75-dot scale with end points labeled "Not at all" and "Completely" accompanied each of the following six stems: "The talk I just heard was persuasive"; "The writer of the speech conveyed sincerity to me"; "The tone of the speech was interesting"; "The typical high school junior would probably consider the speech persuasive"; "The typical high school junior would probably feel the writer of the speech conveyed sincerity"; "The typical high school junior would probably find the speech interesting."

After the six communications had been audited and rated, the subjects filled out a final questionnaire requesting "frankness" and "accuracy." Among the seven questions, the following were pertinent:

> How many times did you attend a place of worship (church, synagogue, etc.) in the past month? Give number______. How many cigarettes did you smoke yesterday? Give number______. How many times did you engage in some form of prayer during the past week? Give number______. On the average how many cigarettes do you smoke in a day? Give number______.

---

[3] We thank D. V. Subashandran for his speech on Indian neutrality.

[4] *Educational Films:* Education Films Which I Enjoy; No conclusions: films are briefly summarized; American speaker. *Cancer Link:* Cancer and Smoking; Conclusion—smoking leads to lung cancer; American speaker. *Indian Neutrality:* Neutralism for Indians; Conclusion—India should remain neutral from both East and West; Native Indian speaker (visiting foreign student). *Christianity is Evil:* Critical Analysis of Organized Christian Religion; Conclusion—Christian teaching is corrupt; American speaker. *No Cancer Link:* Critical Reexamination of Cancer-Smoking Evidence; Conclusion—no proven relationship between smoking and lung cancer; American speaker. *No Draft for Graduates:* College Graduates and the Draft; Conclusion—college graduates should be draft-exempt; American speaker.

There were other comparably personal questions about coffee and gum consumption and preferences for courses. Finally, subjects were instructed:

> Please list below the thoughts and problems that came to you while you were listening and evaluating the speeches. List everything that came to your mind. We are grateful for your frankness, your criticisms, and your helpfulness.

Classification and selection of subjects. Twenty-three undergraduate subjects, 10 males and 13 females, were drawn from introductory summer session psychology courses by telephone. Since a diversity of experimental appointment times could be offered, no refusals were observed and, with two callbacks at most, everyone initially contacted was eventually recruited as a subject. (The same population and recruit procedure were used in Experiments 2, 3, and 4.)

There were 16 nonsmokers and 7 smokers, persons who reported smoking one or more cigarettes "yesterday," and one or more on the average in a day. (Average cigarettes and cigarettes "yesterday" were nearly perfectly correlated in all four experiments.) For smokers, mean daily cigarette consumption was 21.7.

A pilot survey indicated that the present population could be divided naturally, albeit disproportionately, into those who prayed at least once daily and those who prayed less often; into those who attended church every week and those who attended less often. Four or more church attendances per month characterized the high churchgoers ($N = 9$), and eight or more acts of prayer each week characterized the high prayers ($N = 7$). The criterion scores for high church attendance (greater than three per month) and high prayer activity (greater than seven per week) were adopted a priori for each of the four experiments. At the outset it was possible to advance good reasons for believing that churchgoing and praying *would* and *would not* be positively associated. Therefore, both kinds of behavior were used to classify commitment to religion.

Three subjects, of the 26 originally recruited for Experiment I, were omitted in the analysis. One did not follow instructions (by walking about the room, smoking, looking out the window), a second attempted to dismantle the response manipulandum, and a third arrived so late for his appointment that he had to be dismissed.

Summary of procedure. After putting on earphones, the subject read printed instructions (left with him by the experimenter), read the list of speech topics and conclusions (removed by the experimenter before the first speech), and listened to the first speech, pressing the static-removal button ad libitum. The experimenter returned and gave him a rating sheet for the first speech. After rating the first speech, the subject listened to the second speech, and so on. After the six speeches, the subject filled out a questionnaire measuring frequency of religious and smoking behavior. The subjects then responded to the open-ended inquiry about their reactions to the experiment and committed themselves to refrain from discussing their experience with prospective subjects.

Results

Table 1 shows number of static-removing button-press responses, as transformed, to six speeches by 23 undergraduates classified according to smoking behavior, church attendance, prayer frequency, and sex. Each score is a sum of responses for five 15-second segments. An analysis of variance (Winer, 1962,

TABLE 1

STATIC-REMOVING BUTTON-PRESS RESPONSES[a] TO SIX SPEECHES BY LISTENERS CLASSIFIED ACCORDING TO SMOKING ACTIVITY, RELIGIOUS BEHAVIOR, AND SEX: EXPERIMENT I

S classification	Speech topic											
	Education films		Smoking leads to lung cancer		Indian neutrality[b]		Christianity is evil		Smoking does not lead to lung cancer		No draft of college grads	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Smokers (7)	8.5	5.2	11.3	6.3	10.1	6.6	13.7	8.3	17.3	7.8	16.9	7.4
Nonsmokers (16)	9.4	4.4	13.5	7.0	10.8	7.0	12.0	8.3	12.2	6.5	11.6	7.7
Frequent churchgoers (9)	9.1	5.1	12.4	7.3	10.2	7.2	9.6	10.1	12.2	6.5	11.8	7.2
Infrequent churchgoers (14)	9.2	4.3	13.0	6.3	10.8	6.7	14.4	6.4	15.0	7.8	14.1	8.4
Pray frequently (7)	8.7	3.7	11.0	4.6	7.9	4.9	7.7	8.1	9.9	4.9	9.1	7.0
Pray infrequently (16)	9.3	4.6	13.6	7.3	11.7	7.3	14.6	7.6	15.4	7.5	15.0	7.8
Men (10)	9.7	4.8	14.9	7.0	12.8	5.9	14.7	8.9	16.3	7.7	15.5	8.4
Women (13)	8.7	4.4	11.2	6.1	8.9	7.1	10.8	7.5	11.8	6.3	11.4	7.3
Combined (23)	9.1	4.5	12.8	6.6	10.6	6.8	12.5	8.2	13.7	7.2	13.2	7.9

Note.—*N*s in parentheses. See definitions of classifications in text.

[a] After being subjected to a square-root transformation (Mosteller & Bush, 1954, pp. 326–327).

[b] Spoken in the Indian language, Tamil. See text.

p. 112) of the data in the bottom row yielded an F for speech topic that was significant beyond the .01 level. Hence, the communications elicited different amounts of button pressing. If the button-press device measured eagerness to clarify a communication, less pressing should be elicited by the boring Educational Films speech and the incomprehensible India Should be Neutral. Responses to Educational Films were lower than to the other five speeches. This statistically reliable ($p < .01$, Winer, 1962, p. 112) effect could be attributed to "warm-up" and initial diffidence about pressing the button as well as to the aforementioned boring quality of the message. Subjects' postexperimental comments indicated most found the films speech "boring," "not interesting," etc., and many reported being unsure about what they were supposed to do until the beginning of the second speech.

The overall rate of response to the foreign language message (in Tamil) was 10.6, a value reliably lower ($p < .05$) when compared with means of the other messages. Both the Educational Films and Tamil button-press results encouraged interpreting the static-removal behavior as information receptivity.

The hypothesis was that persons will press more frequently to eliminate static from consonant than from dissonant messages. The 7 smokers (four males) were *less* eager—mean number of presses was 11.3—to clarify Smoking Leads to Lung Cancer than the 16 nonsmokers, 13.5 presses; the smokers were *more* receptive to Smoking Does not Lead to Lung Cancer, 17.3, than the nonsmokers, 12.2. The interaction for the smoking (smokers versus nonsmokers) and communication (leads versus does not lead) factors was reliable. An analysis of variance (Winer, 1962, p. 376) yielded an interaction F of 17.38 ($df = 1, 21$), which is significant beyond the .01 level.

A further test of the hypothesis was to examine the static-removal behavior elicited by Christianity is Evil from persons who differed in devotedness. The relevant data are contained in the third to sixth rows of Column 4 in Table 1. Infrequent churchgoers were more eager to clarify the antichurch message than frequent churchgoers. The difference between the mean button-press scores yielded a t of 1.40, which approached reliability. Persons who prayed frequently were less receptive to Christianity is Evil than persons who prayed infrequently; the one-tailed t was significant at the .05 level. The dissonance-avoidance hypothesis thus received some additional, albeit weak, support: high commitment to religious behavior led to lowered receptivity to a communication attacking Christianity.

Feather's (1963, p. 162) contention that perception of relevance importantly governs information receptivity was disputed by Brock (1965, p. 17). In the present study it was assumed that a message about draft abolishment would be perceived as more relevant by men than women. The direction of the difference in means was in accord with the relevance view as can be seen in the seventh and eighth rows of Column 6 of Table 1. Men pressed the button to hear Abolish Draft more than did women. However, the difference did not reach significance at the .10 level, and there was a marked tendency for men to press the button more vigorously, regardless of speech topic. The relevance view might also predict more interest by smokers than nonsmokers in *both* communications about smoking, and more interest by very religious persons in the communication about Christianity. Since both predictions were unsupported, the relevance notion advanced by Feather (1963) did not fit the present data very well.

Discussion

The results of Experiment I, while encouraging, were certainly not decisive. Further verification was required in the light of the following criticisms. First, it could be complained that the obtained results were in some degree an artifact of order of presentation of the communications. Second, comparison of the smokers' with nonsmokers' data in Rows one and two of Table 1 showed smokers slightly more interested in Christianity is Evil and decidedly more interested in Abolish Draft. Perhaps some organismic factor exists which makes smokers more interested in unconventional advocacy and, in the summer of 1964 when the present data were collected, Smoking Does not Lead to Lung Cancer was certainly a minority point of view. In this fashion the interaction between smoking and speech topic could be explained without recourse to dissonance theory. Third, there is the possibility that differences in subjective utility accounted for the results. In terms of an utility interpretation, smokers avoided the Smoking Leads to Lung Cancer message because that message offered no practical assistance, while the opposite message was sought out because it seemed useful in combating a widely held but discrepant belief. Similarly, persons who felt they were neglecting their religious obligations would find it useful to know information which supported their indifference. Other writers (Festinger, 1964, Chapter 4; Freedman, 1965; Freedman & Sears, 1965) have presented reasoning and data to show that perceived usefulness may be a determinant of information receptivity. Fourth, an unforeseen problem arose in the collection of the data. Many subjects "learned" the response-reinforcement contingency, that is, 3 seconds of static-free message for five presses. Once that contingency was mastered the interfering static could be avoided almost automatically. Postexperimental discussion with the subjects confirmed that many listeners concerned themselves in part with making the "correct" number of presses to keep the messages static free. It was encouraging, of course, that interpretable results were obtained in spite of this "learning set" artifact.

EXPERIMENT II

Method

The following modifications were introduced in the procedure and design of Experiment I. In the second, third, and fourth experiments, subjects were randomly assigned to one of four patterns of order of presentation of the communications. The four numerals following the titles of the communications indicate the order of presentation in the first, second, third, and fourth patterns. The first pattern was the same order of presentation used in Experiment I: Educational Films, 1,1,1,1; Smoking Leads to Lung Cancer, 2,5,5,2; India Should be Neutral, 3,3,6,6; Christianity is Evil, 4,4,4,4; Smoking Does not Lead to Lung Cancer, 5,2,2,5; Abolish Draft of College Graduates, 6,6,3,3. The patterns permitted analysis of the possible effects of order of presentation of the pro- and antismoking messages and of early versus late placement of the abolish draft message. Educational Films was retained in first position in each pattern since it was the only message that could be "wasted" while subjects adjusted to the listening situation. The other five messages were involved in tests of the hypotheses and alternative interpretations.[5]

In Experiment II an attempt was made to test a utility interpretation of static-removal behavior. As supplement to the instructions of Experiment I, the experimenter told the subject "In addition to rating the persuasiveness and sincerity of each speech, you will be asked to analyze the content of one of the speeches. You have been assigned one speech for detailed analysis of the content." In the standard list of speech topics given to the subject before listening began, a special instruction was indicated for one of the topics, either Educational Films or Abolish Draft. The special instruction was, "Give an analysis of the content." Finally, a card was left with the subject on which the words "college students and the draft" or "educational films" appeared in the bracketed space. On the card was typed "You have been assigned to write a content analysis of a speech on [ ]. The value of your written analysis will depend upon an accurate understanding of the speech." At the end of the listening session, that is, after all six speeches had been audited and rated, the subjects were given the following form. "Summary of Analysed Speech: Instructions: Write two short sentences summarizing the two most important points in the speech. Be brief."

In Experiment I, 3 seconds of static-free communication were obtained for five button presses. By appropriate rewiring involving a stepper relay, the "correct," that is, static-interrupting, number of button presses was made to vary randomly from one to nine. Thus the subject could not learn a contingency between presses and static elimination.

From the same population used in Experiment I, 34 subjects were drawn, 12 males and 22 females. There were 10 smokers and 24 nonsmokers; mean daily cigarette consumption was 11.8. Six subjects, of the 40 originally recruited for Experiment II, were omitted in the analysis. One was dismissed after he accidentally looked in on the experimenter and saw the two tape recorders, event recorder, etc. Two refused to continue midway in the experiment, one fell asleep during the experiment (she explained she had studied most of the previous night), and another wrote letters to her family instead of listening to the speeches. One arrived so late for his appointment that he had to be dismissed.

---

[5] This admittedly crude approach to controlling order of presentation was used instead of a Latin square because the occurrence of carry-over effects could not be confidently discounted. For example, smokers receiving Smoking Leads to Lung Cancer beforehand might hear Smoking Does not Lead to Lung Cancer with more accumulated dissonance than smokers receiving the no-link message without prior exposure to the disturbing cancer-link message.

Results

A table, not presented here, was composed for Experiment II in the format of Table 1. Comparison of the data from Experiment I and Experiment II showed no overall differences in button-press rate. A statistical test confirmed that the order of magnitude of button pressing was approximately the same in both experiments. Hence, introduction of the randomizing device between button presses and procurement of static-free communication did not alter the volume of responding. Postexperimental questioning confirmed that the unwanted "learning set" artifact of Experiment I had been removed. As in Experiment I, the communications elicited reliably different amounts of button pressing. And again, Educational Films and India Should be Neutral elicited reliably less pressing than the other communications.

To check on the possible effects of order of presentation of the speeches, responsivity to the six communications was considered under each of the four patterns described in the Method section. An appropriate analysis of variance showed that order of presentation did not affect button-press rate. That is, whether Smoking Leads to Lung Cancer and Smoking Does not Lead to Lung Cancer appeared in fifth or second position made no difference. Similarly, early or late presentation of India Should be Neutral and Abolish Draft of College Graduates had no effect. Responsivity to Christianity is Evil, which always appeared in fourth place, was unaffected by preceding patterns of communication order. These negative findings were unsurprising since the procedure required the subject to rate each speech before listening to another speech, and the ratings focused upon speech style rather than content. Thus, the speech(es) that had gone before and the speech(es) to follow probably did not affect the subject's receptivity to the current speech.

The subjects were told that they would have to analyze the content of one of the speeches, either Educational Films or Abolish Draft. It was expected that button pressing to the speech involved in a subsequent task would be elevated in comparison to the rate of pressing to the same speech when it had no subsequent utility. Educational Films was the content-analysis topic for 14 subjects, and their mean button press was 9.8, a value not reliably higher ($p > .50$) than the 8.4 obtained from the control subjects (for whom Abolish Draft was the content analysis topic). Abolish Draft was the analysis speech for 20 subjects; the mean response was 10.5, a value actually lower, but not reliably, than the 13.6 elicited from 14 subjects who were not led to expect that Abolish Draft would be used in a subsequent task. Subjects did not forget which speech they were supposed to analyze; all correctly wrote two pertinent summarizing sentences at the end of the listening session. In sum, there was clear failure to increase button pressing by making subjects believe their careful listening to selected speeches would be useful later on.

Data pertinent to tests of the dissonance-avoidance hypothesis are shown in the first and second columns of Table 2 and Table 3.

TABLE 2

STATIC-REMOVING BUTTON-PRESS RESPONSES[a] TO SMOKING LEADS TO LUNG CANCER (CANCER LINK) AND SMOKING DOES NOT LEAD TO LUNG CANCER (NO LINK) BY LISTENERS CLASSIFIED ACCORDING TO SMOKING ACTIVITY: EXPERIMENTS 2, 3, AND 4[b]

	Experiment 2		Experiment 3		Experiment 4	
	Cancer link	No link	Cancer link	No link	Cancer link	No link
Smokers *M*	11.5	14.8	10.3	14.6	14.7	18.6
SD	3.7	3.3	5.8	2.3	3.7	2.1
N	10		13		9	
Nonsmokers *M*	13.4	12.6	13.7	14.0	18.0	16.8
SD	5.8	8.7	5.1	4.6	2.0	3.5
N	24		15		18	

[a] After being subjected to a square-root transformation (Mosteller & Bush, 1954, pp. 326–327).
[b] See description of experiments in text.

TABLE 3

Static-Removing Button-Press Responses[a] to Christianity is Evil by Listeners Classified According to Church Attendance and Prayer Frequency: Experiments 2, 3, and 4[b]

Listeners	Experiment 2			Experiment 3			Experiment 4		
	M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
Frequent churchgoers	7.3	5.2	6	11.1	5.8	13	15.1	1.7	8
Infrequent churchgoers	12.6	5.8	28	14.2	4.4	15	19.4	1.0	19
Pray frequently	7.0	5.2	8	9.1	6.4	7	15.3	2.2	7
Pray infrequently	13.1	5.0	26	14.0	4.3	21	19.1	1.7	20

[a] After being subjected to a square-root transformation (Mosteller & Bush, 1954, pp. 326–327).
[b] See description in text.

The hypothesis was that persons will press more frequently to eliminate static from consonant than from dissonant messages. The 10 smokers (4 males) pressed *less* to clarify Smoking Leads to Lung Cancer, $M = 11.5$, than the 24 nonsmokers, 13.4; the smokers were *more* receptive to Smoking Does not Lead to Lung Cancer, 14.8, than the nonsmokers, 12.6. A test of the interaction for the smoking (smokers versus nonsmokers) and communication (cancer link versus no link) factors yielded an F of 5.47 ($df = 1, 32$) which is significant at the .05 level. Thus, the interaction of Experiment I was reproduced in Experiment II, although not as strongly. As in Experiment I the interaction pattern was obtained only for the categorization of smokers. In the first three columns of Table 3 it can be seen that frequent churchgoers and frequently praying persons were less receptive to Christianity is Evil than infrequent churchgoers and infrequently praying persons. The one-tailed t's were significant at the .05 and .01 levels, respectively.

Whether persons are more receptive to seemingly relevant communications was checked again. As in Experiment I men pressed the button more than women for all communications including the Abolish Draft message. For the Abolish Draft message, however, the difference in favor of men yielded a t of .4 which, of course, is far from significant. Smokers were not more receptive to both messages about smoking, and persons high on the religiosity attributes were not more eager to clarify the message about religion. Again, Feather's (1963) relevance concept did not obtain support.

In Experiment I it was noted that smokers were not only more receptive than nonsmokers to Smoking Does not Lead to Lung Cancer but also to Christianity is Evil and Abolish Draft. In Experiment II smokers continued to be unreliably more responsive to Christianity is Evil, but this time it was the nonsmokers who pressed more on the average to clarify Abolish Draft. Hence, the earlier suggestion that smokers are attracted to unconventional advocacy appeared less likely.

Discussion

For the samples used in Experiments I and II the proportion of persons classified as smokers was the same, .30. However, in Experiment II reported mean cigarette consumption was about half that in Experiment II. Although this difference was not statistically reliable (two-tailed t-test $p < .20$), it may have been large enough to account for the fact that the Smoking × Communications interaction was not as strong here as it was earlier. Smoking Does not Lead to Lung Cancer and Smoking Leads to Lung Cancer should be more consonant and dissonant, respectively, in proportion to the magnitude of commitment to smoking. Pertinent correlational analyses are presented later.

In sum, order of presentation, relevance, attraction to unconventional positions (novelty), and subjective utility seemed to be having negligible effects in the present paradigm. It was possible, however, that the manipulation of utility expectation was weak. Experiment III accordingly employed a stronger induction of perception of the subsequent usefulness of selected communications.

EXPERIMENT III

Method

Experiment III differed from Experiment II only in the attempt to elicit increased button pressing as a function of manipulated subjective utility. The "content analysis" instruction was continued but the card left with the subject during audition of all speeches was a signed commitment form. The form read "Psychology Research Project: Analysis of Speech. Instructions: below write the topic of the speech you will analyse and sign your name. Topic_________. Signature_________." As in Experiment II, the analysis topic was either "college students and the draft" or "educational films."

From the same population used in Experiments I and II, 28 subjects were drawn, 19 males and 9 females. There were 13 smokers and 15 nonsmokers; mean daily cigarette consumption was 9.9. Twenty-seven subjects, of the 55 originally recruited for Experiment III, were omitted in the analysis. One was dismissed after he accidentally looked in on the experimenter and surveyed the experimental equipment. One subject left the room without explanation after the third speech, and another read a text, instead of listening, and acknowledged she had been studying for an exam. Three subjects were foreign students who did not seem to understand the instructions and rating forms. Twenty-one were omitted because they gave one or fewer button-press responses to two or more communications. Why 21 subjects responded insufficiently in Experiment III turned out to be an unanswerable question. Experiment III was conducted during the hottest part of a hot Iowa summer in unairconditioned facilities, and perhaps the extreme heat produced lassitude and/or affected the sound transmission in some unknown fashion. (Sixteen of the 28 subjects retained in the analysis complained of heat in their postexperimental comment; the number so complaining was negligible in the other three experiments.) In any event, there was no satisfactory explanation.

Results

A table, not presented here, was composed for Experiment III in the format of Table 1. Overall button pressing was of the same magnitude as in Experiments I and II, the communications elicited reliably different amounts of button pressing, and Educational Films and India Should be Neutral elicited reliably less response than the other communications. There were no effects of order of presentation.

Leaving a signed commitment form with the istener surprisingly had no effect on his receptivity to the speech he committed himself to analyze. Educational Films was the emphasized speech for 15 subjects and Abolish Draft, for 13 subjects. Educational Films elicited 9.7 presses when it was emphasized and 11.0 when it was not; Abolish Draft elicited 13.3 when it was emphasized and 11.8 when it was not. These differences did not approach reliability. Again, as in Experiment II, all subjects correctly wrote two summarizing sentences showing that they had not forgotten which speech was to be analysed.

The interaction F for the smoking (smokers versus nonsmokers) and communication (cancer link versus no link) factors was 4.37 ($df = 1{,}26$), which is significant at the .05 level. The means and standard deviations are shown in the third and fourth columns of Table 2. Thus, the dissonance-avoidance hypothesis continued to be reliably, albeit somewhat weakly, supported: smokers were more eager to clarify no link than cancer link and nonsmokers showed a very slight trend in the opposite direction. The results for the Christianity is Evil message, as shown in the fourth, fifth, and sixth columns of Table 3, were also supportive: frequent churchgoers pressed the button less than infrequent churchgoers (one-tailed t-test $p < .10$); frequently praying persons were less receptive than persons who prayed infrequently (t-test $p < .025$).

Males continued to press the button more vigorously than females but the difference in favor of males for the Abolish Draft message did not approach reliability. Other tests of the relevance hypothesis, as earlier described in Experiments 1 and 2, were similarly unconfirmatory here.

It was noted above that data from 21 subjects were excluded because they gave one or fewer button-press responses to two or more communications. Additional analyses, *including* these cases, were performed even though inclusion of these 21 cases led to some nonnormality in the data. Most of the previous results were reobtained: the communications elicited reliably different amounts of button pressing; Educational Films and India Should be Neutral elicited reliably less response than other communications; no effects of the utility emphasis were observed; and the interaction F for the smoking and communication factors was 4.17 ($df = 1{,}47$), significant at the .05 level. However, the effects of churchgoing and prayer upon response to Christianity is Evil did not approach reliability when the 21 low responders were included.

Discussion

The proportion of persons classified as smokers in Experiment III was .46, a higher value than the .30 of the earlier experiments. Here mean cigarette consumption was 9.9, a lower value than the reported amounts of 11.8 in Experiment II and 21.7 in Experiment I. The obtained F for the Smoking × Communications interaction was also lowest here, suggesting again a relationship between degree of behavioral commitment to smoking and differential receptivity to the pertinent communications.

Experiment III repeated the failure of Experiment II to affect static-removing button pressing by manipulation of subjective utility. That persons will be especially receptive to messages which are useful to them is a belief that is both plausible and that is held by other students of selective exposure (e.g., Festinger, 1964, Chapter 4). Since "common sense" and experts appeared in agreement on the utility hypothesis, it seemed necessary to try again to induce differences in expected utility that would measurably affect the dependent variable. In the fourth experiment, expected utility was induced in a more powerful fashion than previously.

Experiment IV was also used to shed light on the generality of the repeated findings in support of the dissonance-avoidance hypothesis. One readily modifiable dimension was the loudness of the message relative to the static input. It seemed worthwhile to show that the present results, even though restricted to a type of operant responding, were not limited to particular decibel values.

EXPERIMENT IV

Method

Experiment IV differed from the earlier experiments in that the volume setting for the spoken messages was reduced from 4.5 to 3.5 with the consequence of reducing the decibel range about two decibels, from about 33 to about 31. This reduction made the static considerably more masking.

In Experiment IV, in order to emphasize the usefulness of one speech, the subject read the following printed instructions. "Supplementary Instructions. In addition to rating the persuasiveness and sincerity of each speech, you will be asked to take a fact quiz on one of the speeches. You have been assigned one speech on which you will be quizzed." On the standard list of speech topics, a sentence accompanied either Educational Films or Abolish Draft: "You will be quizzed on the content of this speech." Further, the subject was given a sample quiz question.

> You have been assigned a fact quiz for the speech on (college students and the draft). How well you do on the quiz will depend on an accurate understanding of the speech. Here is a sample question: Sample Question 1. The speech on college students and the draft was chiefly concerned with ROTC. True *False*.

If Educational Films was the quiz speech, the words "on educational films" appeared in the parentheses and the sample question was "The speech on educational films emphasized jet aircraft. True *False*." Finally, the subject was given a commitment form to sign. "Below write the topic of the speech on which you will be quizzed and sign your name. Topic________ Signature__________." The commitment form was retained by the listener as well as the sample question.

From the same population used earlier 27 subjects were drawn, 21 males and 6 females. There were 8 smokers and 18 nonsmokers; mean daily cigarette consumption was 14.8. Four subjects, of the 31 originally recruited for Experiment 4, were omitted in the analysis because they had information about the experiment prior to their participation.

Results

Increasing the "maskingness" of the static in Experiment IV had a marked effect: button pressing was significantly greater ($p <$ .01) than before in all categories and cells of the data analysis. The significant findings of the earlier studies were replicated; they are listed in the first and third paragraphs of the results for Experiment III above and need not be repeated again here. The data from Experiment IV are shown in Tables 2 and 3. The interaction F for the smoking and communication factors was 10.68 ($df = 1,25$), which is significant at the .01 level. Receptivity to Christianity is Evil was greater from persons low in church attendance ($p < .01$) and in prayer ($p < .01$).

The attempt to increase receptivity to selected speeches by leading subjects to expect a quiz on the speech topic, failed. Educational Films was the quiz speech for 14 subjects and Abolish Draft for 13 subjects. The former speech elicited 13.9 presses when it was a quiz topic and 15.6 when it was not; the latter elicited 18.8 presses when it was a quiz topic and 17.4 when it was not. These differences were unreliable. The relevance-receptivity hypothesis again went unsupported. Females' responsivity to Abolish Draft was unreliably greater than that of the males and the other tests of the relevance notion were negative.

Discussion

The elevation of button pressing in response to reduced loudness of the messages showed that the dependent behavior was aimed at obtaining *clear* communication. *That persons worked more to clarify messages having a priori consonant rather than dissonant face value is a result considered obtainable with different combinations of message and static volume.* Our resources did not permit us to explore limits.

Lowering the loudness of the speeches increased button pressing and also, it seemed, reduced variability. The variances computed for Experiment IV were, on the average, about half those obtained earlier. An entirely satisfactory explanation for this reduction has not occurred to the writers. It was possible that fatigue imposed a ceiling on pressing, and thus a general increase in button pressing would produce a distribution of scores that is foreshortened in its upward tail.

The proportion of persons classified as smokers here was .33, a value close to the .30 obtained in Experiments I and II. It was noted that mean daily cigarette consumption here, 14.8, was less than consumption in Experiment I and more than in Experiments II and III. Similarly, the F value for the Smoking × Communications interaction was second in magnitude. Thus the four experiments yielded a close correspondence between magnitude of the Smoking × Communications interaction and degree of behavioral commitment to smoking. However, this relationship between commitment to behavior and differential receptivity to consonant and dissonant messages appeared to be specific to the smoking issue. Appropriate analyses showed that the differential effect of devotedness on responsivity to Christianity is Evil did not increase with increase in church attendance and prayer frequency.

Freedman and Sears (1965) wrote: "The evidence strongly supports the contention that information which is expected to serve a practical purpose is preferred to less useful information [p. 82]." Although college students presumably will pay attention to information in order to do well on a subsequent quiz, Experiment IV failed to show an effect of perceived utility. Postexperimental interviews showed that subjects expected the quiz and knew the quiz topic. It may still be "true" that persons are especially receptive to useful information but the current failures to demonstrate such receptivity will have to be accounted for in any comprehensive statement of the utility-receptivity doctrine.

OVERALL ANALYSES

Correlations

The above-noted relationship between magnitude of commitment to smoking and strength of the Smoking × Communications interaction was checked further by internal analyses. These correlational analyses employed the entire sample of smokers ($N = 39$) for the four experiments in order to have a sufficient number of cases. The product-moment correlation between number of cigarettes consumed and button pressing to clarify Smoking Leads to Lung Cancer was negative (−.20), but only in the male subgroup ($N = 23$) did the r reach statistical significance ($r = -.47$, two-tailed $p < .05$). The product-moment correlation between smoking and button pressing to clarify Smoking Does not Lead to Lung Cancer was positive for both sexes, but only the combined r approached statistical reliability, $r = +.26$, two-tailed $p < .10$. Hence, for males at least, smoking was inversely related to receptivity to the dissonant cancer-link message and positively related to receptivity to the consonant no-link message. Both findings, of course, supported the hypothesis and corroborated earlier results.

A correlation was computed between church attendance and prayer frequency in each experiment and over the entire series ($N = 112$). In each experiment, the obtained r was not reliably greater than zero and the overall r, $+.23$, though statistically significant with $df = 110$, was considered insubstantial enough to justify the a priori decision to treat church attendance and prayer as separate indexes of religious commitment. In each of the four experiments the correlation between church attendance and prayer frequency, on the one hand, and receptivity to Christianity is Evil, on the other, was negative. None of these correlations, however, attained the .05 level of reliability. The overall r, with $df = 110$, between church attendance and receptivity to Christianity is Evil was $-.20$, a value reliable at the .05 level; the overall r between prayer frequency and receptivity to Christianity is Evil was $-.28$, reliable at the .01 level. These results supported the dissonance-avoidance hypothesis.

Analyses of Variance

The results for Experiment I (Table 1) showed an overall difference between men and women in amount of button pressing, and that men approached the cancer issue differentially. The overall correlational analysis suggested that the dissonance-avoidance effect was restricted to male smokers. Furthermore, the experiments that included the greatest proportion of males (Experiments I and IV) had the heaviest smokers and showed the strongest interaction between the smoking and communication factors. Thus, in the present series of experiments, sex and extent of smoking may have been confounded in their effect on button pressing. Sex and smoking may also have been confounded with church attendance and prayer.

TABLE 4

Summary of Analysis of Variance of Button-Press Responses to Smoking Leads to Lung Cancer (Cancer Link) and Smoking Does Not Lead to Lung Cancer (No Link) by Listeners Classified According to Sex and Smoking Activity: 96 Subjects Selected[a] from All Four Experiments

	df	*MS*	*F*
Between			
Smokers versus non-Smokers (A)	1	11.51	<1
Males versus females (B)	1	807.41	16.28**
A × B	1	.36	<1
Error	92	49.59	
Within			
Cancer link versus no link (C)	1	48.51	5.98*
A × C	1	334.61	41.21**
B × C	1	.81	<1
A × B × C	1	9.63	1.19
Error	92	8.12	

[a] See explanation in text.
* $p < .05$.
** $p < .001$.

In the first of several overall ($N = 112$) analyses of variance (unweighted-means solution), experiment, smoking, and sex were between-subject factors, and communication topic (Smoking Leads to Lung Cancer versus Smoking does Not Lead to Lung Cancer) was a within-subject factor. There was a reliable ($p < .01$) main effect of sex but sex did not enter into any interactions that approached reliability. There was no main effect of smoking but the Smoking × Communication interaction yielded an F of 20.2, significant beyond the .001 level. Further overall analyses were carried out in which sex, smoking, experiment, prayer, and church attendance were the factors, and response to Christianity is Evil was the dependent datum. In these analyses sex and smoking did not enter into interactions that approached reliability. These results weighed against the possible confoundings discussed above.

A final overall analysis of variance (least-squares solution) was carried out with a matched selection of 96 subjects. There were 16 female smokers, and 16 male smokers were selected from the pool of 23 male smokers so that daily cigarette consumption for the male and female groups was approximately matched. A random drawing of 32 males and 32 females from the nonsmokers completed the selection. Smokers pressed the button more than nonsmokers to hear Smoking Does not Lead to Lung Cancer and less to hear Smoking Leads to Lung Cancer. The interaction was highly reliable as shown in Table 4. Table 4 also shows a reliable main effect of sex but, again, no interactions. Thus, support for the dissonance hypothesis was obtained with sex effectively partitioned out.

General Discussion

The present series of four experiments repeatedly yielded relationships between behavioral commitment and behavioral self-exposure to pertinent propaganda. Smokers were more likely than nonsmokers to attempt removal of static from a message disputing the link between smoking and lung cancer; smokers were less likely than nonsmokers to attempt to clarify a message affirming a smoking-cancer link. The greater the amount of cigarettes smoked, the stronger these relationships appeared to be. In each of the four experiments smokers were more receptive to supportive information (denial of the smoking-cancer link) than to nonsupportive information (affirmation of the smoking-cancer link). In another area, religious behavior, persons strongly committed to religion, as evidenced by their greater recourse to prayer and church attendance, were less eager to clarify a message attacking Christianity than were persons weakly committed to religion. These principal findings were considered solidly replicable; they did not seem explainable by mechanisms other than dissonance avoidance. Straightforward derivations from a relevance position were unconfirmed; the role of novelty in producing the obtained pattern of receptivity scores could be empirically discounted, as discussed below; and three attempts to affect receptivity by varying information utility failed. In the present paradigm, therefore, the bulk of the variance seemed accountable for in terms of individual differences in response rate and message attributes such as boringness, intelligibility, and consonance.

The dependent variable, message-clarifying button pressing, was considered a valid measure of information receptivity. A question may be raised, however, from consideration of the failure to support the utility hypothesis. Perhaps attention to button pressing interfered with attention to the messages. One consequence of this "competing response" interpretation is that button pressing should be lower for the emphasized speech topic and higher for the nonemphasized topic. The previously described results, however, did not provide any consistent support for a "competing response" view. Merely *clear* discourse, in the case of the Tamil message, was insufficient to produce elevated responding relative to the other messages, and merely *intelligible* discourse, the Educational Films communication, did not elicit increased responding. Button pressing apparently increased when the message was both meaningful and supportive.

The findings may be contrasted with previous failures systematically to confirm the dissonance-avoidance hypothesis (Freedman & Sears, 1965). Comparison of the present experiments with an earlier one by Brock (1965) is instructive because Brock used the same subject population and message stimuli. His subjects indicated their reading preferences among magazine articles. The dependent measure was a preference ranking for the titles "Smoking Leads to Lung Cancer" or "Smoking Does not Lead to Lung Cancer." These topics, of course, were the same ones used in the present series. In the earlier experiment (Brock, 1965), smokers did not differ reliably from nonsmokers in their preference for the message linking smoking to lung cancer, that is, there was no evidence for avoidance of nonsupportive information. The noteworthy difference between the current series and Brock (1965) is that subjects in the earlier study were evaluating information prospectively while here, confrontation by the discrepant information was already underway. The implication is that present, past, and future time contexts should be distinguished in subsequent work on information receptivity. Perhaps selective exposure, attending to the message at all versus not at all, must be distinguished from selective *attention*, exposure to the message with attention only to parts. When the subject is actually exposed to information, dissonance avoidance may be readily demonstrable (present series; Cohen, Brehm, & Latané, 1959; Festinger, 1957, pp. 162–176; Jecker, 1964, pp. 76–77).[6]

---

[6] Studies failing to show avoidance of concurrently presented dissonant material seem to be ones in which the prior commitment or decision was clearly uninvolving and/or in which the dependent measure was somewhat unrefined. For example, Sears and Freedman (1965) employed a mock-jury situation; their experimenter ran four to five subjects at a time and the dependent measure was the time intervening between the experimenter's departure from the subject's cubicle and the subject's signal that he had finished reading the material left with him by the experimenter. Some Sears-Freedman subjects could have daydreamed and read little before signalling the experimenter while others could

For the case of retrospective information receptivity, wherein the subject is considering previous acquired information, avoidance of dissonant information has been sufficiently documented (see references in Brock, 1965, p. 16; Jecker, 1964, pp. 78–80). Difficulty in confirming the dissonance-avoidance hypothesis seems restricted to the case in which the subject is given a modicum of information, for example a title, regarding messages to which he might be exposed *in the future*. It is for this case only that dissonance theory appears inadequate; additional processes have to be postulated. The reformulation offered by Brock (1965, pp. 17–18) is the only one so far available that encompasses both positive and negative evidence and also allows for the derivation of other models as special cases. The core of Brock's (1965) statement was:

> (*a*) persons will expose themselves to cognition dissonant with their choice (of behavior, opinion, object, etc.) in direct proportion to the number and importance of prior cognitions consonant with the choice; (*b*) the gratification of dissonance reduction is proportional to the magnitude of dissonance arousal; (*c*) when dissonant cognition is not avoided, receptivity to it will be proportional to the magnitude of the anticipated discrepancy [p. 17].

These propositions are general; they permit the derivation of specific determinants of receptivity as particular instances. For example, Freedman (1965), Festinger (as cited by Freedman & Sears, 1965, p. 76) and Canon (1964) have speculated about the role of opinion confidence in determining self-exposure to nonsupportive information; Mills (1965b) similarly employed certainty (that a choice is correct) as the principal determinant of receptivity to dissonant information. Opinion confidence and choice certainty are particular kinds of consonant preexposure cognitions; according to Brock's (1965) proposition *a*, as such cognitions increase in number and salience, exposure to dissonant cognitions will increase. Proposition *a* not only handles well the confidence (Freedman, 1965) and certainty (Mills, 1965b) views, but it may be the most parsimonious account of the effect of one-sided versus two-sided presentations upon subsequent information selectivity in adversary situations (Sears, 1965). Sears' conclusion can be considered a paraphrase of proposition *a:*

> the available evidence supports the notion that preferences for counter-information follow when a person has had a one-sided past history of exposure to information on an issue; i.e., when considerably more of the information he has seen favors one side than the other. This seems to hold even when the one-sidedness is due to biased information content rather than to differential exposure to spokesmen for each side [Sears, 1965, p. 373].

During approximately the same period in which selective exposure has come to preoccupy social psychologists, there has been a far more rapid growth in a seemingly relevant experimental area, the study of curiosity and exploration (Berlyne, 1966; Fowler, 1965). In these studies the stimuli have usually been visual patterns or varieties of stimulus change and the respondents have been diverse—chicks, rats, neonates, and college sophomores. However, messages of the kind that affect attitudes and influence personal decisions and interpersonal behavior have never been employed as effective stimuli by the investigators of exploratory behavior. The many other interesting contrasts and parallels between the selective-exposure and exploratory-behavior literatures cannot be dealt with here. Two observations are worth conveying, however. First, comparison of the literatures, especially including the present results, illumines the distinction between dissonance on the one hand, and incongruity, complexity, irregularity, or heterogeneity on the other. The latter seem to elicit increased responding, whereas it has been shown here that confrontation by a dissonant stimulus depresses responding which would clarify the stimulus. Dissonance results for the individual from the violation of what is psychologically implied by his prior commitment or decision, whereas incongruity, complexity, magnitude of stimulus change, and the like, can be defined without reference to a particular recipient organism. The fact that concurrent dissonance depresses stimulus-clarifying responses while

---

have read and reread the material many times. The two types of behavior could misleadingly yield identical time scores. This kind of artifact was recognized and removed by Jecker (1964) in the analysis of his previously cited supportive results.

incongruity and moderate stimulus change increase stimulus-clarifying responses compellingly establishes the distinction.

The second observation concerns novelty and receptivity. The present experiments suggested that novel stimulation does not necessarily increase stimulus-clarifying responding. Some analysis of the novelty-receptivity hypothesis was presented earlier; it may be further noted that the Tamil message was certainly novel (also "irregular" and "complex"), but receptivity to India Should be Neutral was depressed compared to receptivity to the other messages. A novelty or stimulus change formulation might have predicted increased receptivity to Educational Films because each recited film title was new and different from the previous title; redundancy was much lower than in the communications on smoking, religion, and college draft. It seemed the present paradigm did not evoke the kinds of processes highlighted by the epistemic behavior theorists. Hence, the possibility can be entertained that the generalizations developed by these theorists do not apply to lifelike situations wherein individuals encounter and selectively expose themselves to propagandistic messages. The present static removal technology could facilitate direct tests pitting dissonance avoidance against a boredom drive (Fowler, 1965) or a curiosity motive (Berlyne, 1966).

There have been too few attempts to cast social-psychological issues in the explicit language of S-R theory (e.g., Weiss, Rawson, & Pasamanick, 1963). Integration of a social psychology theory with behavioral theory requires recognition of interrelationships between the theories and *the adoption of common dependent variables,* a common response.

> The classical theories of Hull and Spence define the response in molar terms like pressing a bar, running down an alley, etc., and the quantitative characteristics of a response are predicted from a single terminal construct, excitatory potential. Independent variables such as practice, drive, reward, etc., converge upon excitatory potential and this single number [magnitude of excitatory potential] is used to determine whether the response will occur, how quickly it will occur, and what its amplitude will be [Logan, 1959, p. 297].

The present experiments encourage the use of a specific molar response, static-removing button pressing, in situations for which testable hypotheses can be derived from both behavioral and social-psychological theories. One interesting next step may be the attempt to coordinate and/or contrast S-R and dissonance theories by determining the consequences of the intertranslation of key concepts. This task can be adumbrated here by illustrations from the kind of conceptual exercise that may yield new laws and specification of the limits to extant generalizations. The constructions immediately following depend upon several unstated assumptions, such as that supportive information is rewarding.

Excitatory potential may be read as receptivity, measured by message-clarifying button pressing. There is already good justification (Brock, 1963, p. 329) for the intersubstitutability of dissonance and drive. "Incentive" may be thought of as the magnitude of consonance conveyed in the stimulus message. The greater the drive or dissonance (heavier commitment to discrepant behavior, e.g., smoking; increased personal importance of a prior decision) and the greater the incentive or consonance (messages can be written and scaled to vary in supportiveness for a prior commitment or choice), the greater will be the excitatory potential or rate of message clarification. Interesting empirical questions emerge. Which combinatorial rule for drive and incentive, addition or multiplication, is applicable to an individual selectively attuned to meaningful messages? Is the Hull-Spence approach limited to nonlanguage situations (Logan, 1959, p. 302)? Can the derivation of dissonance avoidance and other selectivity phenomena be pursued within a Hullian-type framework? The experimental paradigm and dependent variable introduced in this report now make such questions answerable.

REFERENCES

AMERICAN CANCER SOCIETY. *Cigarette smoking and cancer.* New York: American Cancer Society, 1963.

BERKMAN, A. *Now and after.* New York: Vanguard, 1920.

BERLYNE, D. E. Curiosity and exploration. *Science,* 1966, **153**, 25–33.

BROCK, T. C. Cognitive restructuring and attitude change. *Journal of Abnormal and Social Psychology,* 1962, **64**, 264–271.

BROCK, T. C. Effects of prior dishonesty on postdecision dissonance. *Journal of Abnormal and Social Psychology,* 1963, **66**, 325–331.

BROCK, T. C. Commitment to exposure as a determinant of information receptivity. *Journal of Personality and Social Psychology,* 1965, **2**, 10–20.

BURES, D. Should college graduates be drafted? No. *The Intercollegian,* 1963, **81**, 5–7.

CANON, L. K. Self-confidence and selective exposure to information. In L. Festinger (Ed.), *Conflict, decision, and dissonance.* Stanford: Stanford University Press, 1964. Pp. 83–95.

COHEN, A. R., BREHM, J. W., & LATANÉ, B. Choice of strategy and voluntary exposure to information under public and private conditions. *Journal of Personality,* 1959, **27**, 63–73.

EHRLICH, D., GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision exposure to relevant information. *Journal of Abnormal and Social Psychology,* 1957, **54**, 98–102.

FEATHER, N. T. Cognitive dissonance, sensitivity, and evaluation. *Journal of Abnormal and Social Psychology,* 1963, **66**, 157–163.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

FESTINGER, L. *Conflict, decision, and dissonance.* Stanford: Stanford University Press, 1964.

FOWLER, H. *Curiosity and exploratory behavior.* New York: Macmillan, 1965.

FREEDMAN, J. L. Preference for dissonant information. *Journal of Personality and Social Psychology,* 1965, **2**, 287–289.

FREEDMAN, J. L., & SEARS, D. O. Selective exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* Vol. 2. New York: Academic Press, 1965. Pp. 58–97.

HOVLAND, C. I. Reconciling conflicting results derived from experimental and survey studies of attitude change. *American Psychologist,* 1959, **14**, 8–17.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale University Press, 1953.

JECKER, J. D. Selective exposure to new information. In L. Festinger (Ed.), *Conflict, decision, and dissonance.* Stanford: Stanford University Press, 1964. Pp. 65–81.

LINDSLEY, O. R. A behavioral measure of television viewing. *Journal of Advertising Research,* 1962, 2, 2–12.

LOGAN, F. A. The Hull-Spence approach. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959. Pp. 293–358.

MILLS, J. Avoidance of dissonant information. *Journal of Personality and Social Psychology,* 1965, 2, 589–593. (a)

MILLS, J. Effect of certainty about a decision upon postdecision exposure to consonant and dissonant information. *Journal of Personality and Social Psychology,* 1965, **2**, 749–752. (b)

MILLS, J., ARONSON, E., & ROBINSON, H. Selectivity in exposure to information. *Journal of Abnormal and Social Psychology,* 1959, **59**, 250–253.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 1. *Theory and method.* Cambridge: Addison-Wesley, 1954. Pp. 289–334.

NORTHRUP, E. *Science looks at smoking.* New York: Coward-McCann, 1957.

SANFORD, T. By their own free choice. *Vital Speeches of the Day,* 1964, **30**, 463–467.

SEARS, D. O. Biased indoctrination and selectivity of exposure to new information. *Sociometry,* 1965, 28, 363–376.

SEARS, D. O., & FREEDMAN, J. L. Effects of expected familiarity with arguments upon opinion change and selective exposure. *Journal of Personality and Social Psychology,* 1965, **2**, 420–426.

WEBB, E. J., CAMPBELL, D. T., SCHWARTZ, R. D., & SECHREST, L. *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago: Rand McNally, 1966.

WEISS, R. F., RAWSON, H. E., & PASAMANICK, B. Argument strength, delay of argument, and anxiety in the "conditioning" and "selective learning" of attitudes. *Journal of Abnormal and Social Psychology,* 1963, **67**, 157–165.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received October 3, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 429–434

SOCIAL EXPERIENCE AND SOCIAL SCHEMATA[1]

LAURA WEINSTEIN

George Peabody College for Teachers

On the assumption that positive connotations about human relationships are learned in a context of positive social experience, it was hypothesized that acceptance by parents and peers would be associated with positive social schemata as measured by Kuethe's felt figure-placement technique. Elementary-school boys were asked to accurately reproduce from memory distances between pairs of figures. As hypothesized, those who underestimated distances between human pairs felt more accepted by their parents. Similar relationships with perceived or actual acceptance by classmates did not hold. An additional finding, that firstborns (higher need for affiliation) are more attracted than later borns to children having positive schemata, was interpreted as suggesting that attributes of the chooser may interact with schema of the chosen in determining sociometric choice.

Kuethe and his associates (Kuethe, 1962a, 1962b, 1964; Kuethe & Stricker, 1963), studying how subjects naturally group human figures, found many consistent patterns. For example, human figures were placed together on a field and were not separated by geometric figures, mixed-sex rather than same-sex figures were paired, and a child was placed closer to a mother than a father figure. These patterns were reflected also in judgment errors made in reproducing distances: subjects underestimated distances between pairs commonly grouped together.

Kuethe accounted for *intraindividual* consistency in these behaviors by positing the existence of social schemata. These schemata are sets of meanings about human relationships which are learned through social experience and which mediate social behavior. His figure-placement techniques provide indexes of the underlying schemata. *Interindividual* consistency is accounted for by commonality of experience, which leads to the learning of similar, shared schemata. A two-step chain is postulated: similar patterns of interpersonal experience lead to similar schemata which, in turn, lead to similar behavior.

The finding by Kuethe and Weingartner (1964) that homosexuals pair same-sex figures while nonhomosexuals pair mixed-sex figures supported the hypothesis that schemata and behavior are related. A study by Weinstein (1965) offered inferential support for the relationship between social experience and schemata. The study was suggested by Kuethe's report that normal adults characteristically underestimate the distance between human figures. Presumably emotional distance is translated into physical distance in the reconstruction task, and the constant error made by Kuethe's adults reflected a social schema, learned through rewarding associations with others, in which people are seen as closely and positively related. On the assumption that emotionally disturbed children experience rejection and punishment in their social relationships, Weinstein hypothesized that these children would show an opposite constant error, that is, that they would overrather than underestimate the distance between human pairs. This hypothesis was confirmed; disturbed children overestimated the distance between human pairs, while normal children as a group made no consistent error.

The present research tests the assumption of a relationship between schemata and social experience more directly by relating measures of the normal child's experience of interpersonal acceptance to his behavior on the social schema measure. Three experiments were performed, primarily to provide cross-validation of results, but also to pursue questions raised as the research progressed. The following hypotheses were tested in Experiment I:

---

[1] This study was supported by Public Health Service Grant MH 929 from the National Institute of Mental Health. The cooperation of the Nashville-Davidson County Metropolitan School System is gratefully acknowledged. Special thanks are due Evelyn Cromwell for her help in collecting the data.

Hypothesis 1. Children who feel accepted by their parents hold a positive schema concerning human relationships (i.e., underestimate distances between human pairs).

Hypothesis 2. Children who are accepted by their peers hold a positive schema concerning human relationships.

EXPERIMENT I

Method

Subjects. The sample consisted of 53 second- through sixth-grade white boys from a lower-middle-class public school. Mean age was 117.6 months ($SD = 18.0$), mean IQ (Lorge-Thorndike Intelligence Tests) was 100.1 ($SD = 14.3$). All subjects had been assessed by their teachers as being within the normal range with respect to behavior. (Academic achievement was disregarded for the purposes of this assessment.)

Measures. Kuethe's reconstruction task, introduced to the children as a study of accuracy in judging distances, was used to index the social schema. The subject looked for 5 seconds at a pair of white felt figures placed 15 inches apart on a large piece of black felt 10 feet in front of him. He was told that when the figures were taken down and given to him he was to replace them exactly as far apart as they had been. Each subject replaced two pairs of figures (two rectangles, a man and a woman), with order of presentation counterbalanced over subjects. Replacement of the rectangles was used as a benchmark in categorizing each subject as overestimating or underestimating the distance between the human figures. If the human figures were replaced closer together than the rectangles, the subject was categorized as underestimating the distance (positive schema); if the rectangles were replaced closer, he was classified as overestimating (negative schema).

The parental acceptance measure assessed the extent to which the child saw himself as meeting his parents' expectations. It was thus a measure of parental acceptance as experienced by the child. A 12-item scale was constructed. The items asked about skill in playing ball, grades and behavior in school, getting angry easily, being good-looking, wishing to be younger or a girl, etc. Each described a child ("Johnny gets good grades in school") and was followed by three questions: "How much like Johnny are you?"; "How much like Johnny does your mother want you to be?"; "How much like Johnny does your father want you to be?". Response alternatives were "not at all," "not much," "pretty much," and "very much." For scoring purposes, these alternatives were assigned numerical scores from 0 to 3. Discrepancy between self-concept and maternal (paternal) expectations was the sum over items of the absolute difference between the alternative chosen in response to "How much like Johnny are you?" and "How much like Johnny does your mother [father] want you to be?" Since discrepancy with the expectations of the mother correlated above .90 with discrepancy with the expectations of the father, the two scores were summed, and analyses are reported for the combined parental discrepancy score. A low discrepancy score indicates high acceptance. Test-retest reliability for this score over an 8-month interval was .64 for a group of 41 subjects.

A sociometric questionnaire, filled out by all children in each subject's class, was the measure of peer acceptance. The questionnaire read: "You are going to have a birthday party. Name five friends that you would ask to your party. Please ask your friends from your classroom. Write the names on the lines." Five lines were provided, and the children were encouraged to list five names. Score was the number of nominations received. Test-retest reliability for this measure with a 4-month interval was .75 for 56 subjects.

Procedure. The social schema and parental acceptance measures were given individually, 1 week apart. Items were read to the children, who answered orally. The parental acceptance measure was not administered to three children who had no father or siblings in the home. The sociometric questionnaires were group administered by the classroom teachers to their classes approximately 1 month later. Since one teacher was unwilling to do this, sociometric data were available for only 42 subjects. To check for possible relationships between age or intelligence and the schema or acceptance measures, group intelligence scores and birth dates were obtained from the school records.

Results

It can be seen from Table 1 that both hypotheses were supported in this experiment. A positive schema was associated with high parental acceptance ($r = .30$, $p < .025$) and with high peer acceptance ($r = .26$, $p < .05$). The acceptance measures were not significantly related to each other (see Table 2). No correlation between age or intelligence and the other measures reached significance. Since none of them exceeded .09, they are not individually presented in the tables.

EXPERIMENT II

The second experiment was designed to cross-validate the schema-acceptance relationships. In addition, a perceived peer-acceptance measure was included to allow more valid comparison of the relationships of parental and peer acceptance to social schema.

TABLE 1

POINT-BISERIAL CORRELATIONS BETWEEN SOCIAL SCHEMA AND ACCEPTANCE MEASURES

Experiment	N	Perceived parental acceptance	Perceived peer acceptance	Actual peer acceptance
I	50	.30**		.26*
II	57	.28**	.02	−.19
III				
Group A	56		.02	−.16
Group B	44		−.07	−.01

Note.—A positive correlation indicates that a positive schema is associated with high acceptance.

* $p < .05$, one-tailed test. (Only 42 subjects were available for this comparison.)

** $p < .025$, one-tailed test.

"Social experience" may refer either to the external situation as objectively assessed or to that situation as perceived and felt by the child. While the two are undoubtedly related, they are not identical, and it seems likely that social schemata are more closely linked with the latter than the former. The parental acceptance items used in Experiment I assess perceived parental acceptance; the sociometric questionnaire assesses actual peer acceptance. In the second experiment, measures of both perceived and actual peer acceptance were obtained and hypothesized to be related to social schema as specified in Hypothesis 2. Hypothesis 1 was tested as in the earlier experiment.

Method

Subjects. The sample consisted of 57 fourth-grade boys from a second school of comparable socioeconomic background. They were screened as in Experiment I. All fourth-grade boys in the school who passed the screening and were present for testing were included in the sample.

Measures and procedures. The measures and procedures utilized in Experiment II were repeated except that the sociometric questionnaire was administered by a research assistant rather than the classroom teacher. Immediately after filling out the sociometric questionnaire, the child was asked to estimate how many children in the class had written his name down. The fact that each child could only have written five names was stressed, and the task was explained as an attempt to find out how accurately children could make the guess. The child's estimate was the measure of perceived peer acceptance. Test-retest reliability for the estimate after a 4-month interval was .49 for 56 subjects. This reliability is rather low, but was considered sufficient for exploratory purposes.

Results

It can be seen from Table 1 that the first hypothesis received additional support in Experiment II: holding a positive schema was again related to perceived parental acceptance ($r = .28$, $p < .025$). In contrast to Experiment I, actual peer acceptance failed to relate to schema in this sample. In fact, the direction of the correlation in this experiment was contrary to hypothesis. Nor did perceived peer acceptance show any relation to social schema. Table 2 indicates that the two peer-acceptance measures were related to each other ($r = .34$, $p < .005$), and that neither was significantly related to the parental acceptance measure.

EXPERIMENT III

Because of the inconsistent results of the first two experiments regarding the relation between schema and peer acceptance, Experiment III was undertaken to provide further tests of the relationship. Experiment III also provided an opportunity to look at the relationship between schema and sociometric nominations in a different way.

Up to this point, the focus of interest had been on the impact of social experience (and particularly social acceptance) on social schemata. It is possible that social schemata may also affect social experience. For example, the behaviors associated with his schemata may affect the child's acceptability to his parents. While this possibility must be taken into account in evaluating the signifi-

TABLE 2

INTERCORRELATIONS AMONG THE ACCEPTANCE MEASURES

	Perceived parental acceptance		Perceived peer acceptance		
				Exp. III	
	Exp. I	Exp. II	Exp. II	Group A	Group B
Perceived peer acceptance		−.10			
Actual peer acceptance	.04	−.01	.34***	.48***	.24

*** $p < .01$, two-tailed test.

cant correlation between schema and parental acceptance, the fact that the parental relationship temporally precedes the formation of schemata permits some confidence that the correlation reflects, at least in part, the effects of parental acceptance on the development of schemata.

The situation with respect to actual peer acceptance is even less clear-cut. The child presumably comes to school with already-formed schemata. A significant correlation between schemata and peer acceptance might reflect the modifiability of the early-formed schemata by later experience. It might reflect instead, or in addition, the effects of already-formed schemata on the acceptability of the child to his peers.

The results of the second experiment suggested, however, that there may not be a relationship between peer acceptance and the schema studied. If further tests of the relationship yielded nonsignificant results, in line with the above reasoning two conclusions would have to be considered: (*a*) Early-formed schemata are not affected by peer acceptance; and (*b*) the schemata and their associated behaviors are not salient for children in making sociometric choices.

The third experiment was designed to explore yet another possibility, an alternative to the second stated above. The schemata and their associated behaviors may be salient, but some children may be attracted to those who hold positive schemata, while others are attracted to those holding negative schemata.[2] Positive and negative schemata may be associated with different interpersonal modes to which children are differentially attracted depending on their own personal characteristics and interpersonal needs. Considering only total sociometric nominations received would mask this process.

An attempt was made in Experiment III to test for a relationship between attributes of the chooser and schema of the chosen by specifying a kind of child who would choose (be attracted to) those holding positive schemata. Firstborn children have been shown to have greater needs for affiliation (e.g., see Schachter, 1959). Presumably their affiliative needs would be more adequately filled in interactions with those who view human relationships as close and supportive (i.e., those who hold positive schemata). The specific hypothesis tested was:

Hypothesis 3. Firstborn children are more likely than later borns to choose children holding positive schemata.

One further possibility was explored in attempting to find a relationship between attributes of the chooser and schema of the chosen—the possibility that the schema the child himself holds may make children holding one or the other schema more attractive to him. The "similarity" principle would suggest that children would choose others with schemata like their own; the "opposites attract" principle would suggest that they would choose those with schemata opposite to their own. The nondirectional hypothesis tested was:

Hypothesis 4. The schema of the child chosen will depend on the schema held by the chooser.

Method

Subjects. Two samples were drawn from a third school, comparable in socioeconomic background to the schools used earlier. The first, designated as Group A, consisted of 56 third-grade boys. The second, Group B, consisted of 44 fifth-grade boys. Children were selected for the samples by the same criteria used in Experiment I.

Measures. Schema and peer-acceptance measures were administered as in Experiment II. However, the schema measure was given to all members of each class so that count could be taken of the number of children holding positive schemata each child chose. Information on sibling status was obtained from the children.

Analysis. To further test Hypothesis 2, point-biserial correlations between schema and the perceived and actual peer-acceptance measures were computed as in Experiment II. Because of the limited range of possible scores on the schema-choice measure (number of children, of the five chosen, who held positive schemata), Mann-Whitney *U* tests were used to test Hypotheses 3 and 4. To test Hypothesis 3, number of children chosen holding positive schemata was tabulated against sibling status (firstborn or later born) of the chooser. To test Hypothesis 4, number of children chosen holding positive schemata was tabulated against schema (positive or negative) held by the chooser.

---

[2] We have thought of children with positive schemata as more warm and open and children with negative schemata as more defensive and alert to threat. This distinction, partially supported by unpublished data indicating that a negative schema is associated with greater defensiveness, had led us to expect positive schemata to be more interpersonally attractive. Negative schemata may also be associated with greater interpersonal effectiveness in the Machiavellian sense, however, and it may be the latter to which some children react positively.

Results

It can be seen from Table 1 that no significant relation was found in these samples between schema and peer acceptance. Hypothesis 2, for both actual and perceived peer acceptance, can be regarded as unsupported by the present research. Hypothesis 4 also received no support. Schema of the chosen was not related to schema of the chooser.

Table 3 shows the data pertinent to Hypothesis 3. (There were eight children without siblings; these are omitted.) The Mann-Whitney *U* test (corrected for ties) yielded a z of 1.70 ($p < .05$) for Group A, a z of 1.52 ($p < .07$) for Group B. The test for the significance of a pair of results (Edwards, 1954, p. 391) gave a χ^2 of 11.67 ($df = 4$, $p < .02$). These results support the specific hypothesis that firstborns are more likely than later borns to choose children holding positive schemata. They also offer support for the more general hypothesis that schemata are salient for sociometric choice, but that the extent to which a particular schema attracts a particular child depends on characteristics of the latter.

TABLE 3

SCHEMA CHOICE AS A FUNCTION OF BIRTH ORDER

	No. children chosen having a positive schema						
	0	1	2	3	4	5	*N*
Group A							
Firstborns	0	2	12	5	3	1	23
Later borns	1	6	13	6	1	0	27
Group B							
Firstborns	0	0	4	3	2	1	10
Later borns	1	5	12	10	4	0	32

DISCUSSION

The importance of the child's relationship with his parents as the prototype of later social relationships has been a major motif in the developmental literature. Piaget (1951, p. 207) has noted:

> Day to day observation and psychoanalytic experience show that the first personal schemes are afterward generalized and applied to many people. According as the first interindividual experiences of the child who is just learning to speak are connected with a father who is understanding or dominating, loving or cruel, etc., the child will tend (even throughout life if these relationships have influenced his whole youth) to assimilate all other individuals to his father scheme.

Similarly, Dollard and Miller (1950) in discussing the therapeutic relationship have pointed out that the patient generalizes to the therapist the same responses which were learned in interaction with authority figures in early life. The present study confirms the importance of this primary relationship in forming attitudes which undergird other social behaviors.

It is generally felt, however, that peer relationships become increasingly important after the child enters school. Do these relationships modify early schemata? The present findings, if taken at face value, indicate they do not. Neither the child's estimate of his popularity with his classmates, nor his actual popularity with them, was related to the schema studied. One problem in the present research, however, is that it deals with peer acceptance in terms of magnitude rather than depth. It selected a group, classmates, which might or might not be focal for the child. It measured peer acceptance in terms of number of nominations from this group, taking no account of the possibility that one very important peer relationship—in or out of the classmate group—might be more potent than many sociometric nominations. A second issue is that it may not be until adolescence, when ties with parents are looser and peer relationships more intense, that relationships with age mates can effectively modify early schemata. Or, Freud, Piaget, and others may be right. Once established through interaction with parents, schemata may be relatively resistant to change.

One reason early-formed social schemata may remain unmodified is that they may operate to structure later social experience. The finding that those with positive schemata are more likely to be chosen by firstborn children (presumably those with high affiliative needs) suggests a way in which the structuring process operates. The behavior flowing from a particular social schema is differentially attractive to persons with differing personality characteristics. The principle underlying the attraction seems to be one of complementarity. Others seek out those whose behavior meets their particular needs, and in maintaining the relationship are likely to continue to expect and to reinforce this behavior. Thus the selection process which determines the composition of free groups may function to confirm the child in the schemata he has already developed.

REFERENCES

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

EDWARDS, A. L. *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

KUETHE, J. L. Social schemas. *Journal of Abnormal and Social Psychology,* 1962, **64**, 31–38. (a)

KUETHE, J. L. Social schemas and the reconstruction of social object displays from memory. *Journal of Abnormal and Social Psychology,* 1962, 65, 71–74. (b)

KUETHE, J. L. The pervasive influence of social schemata. *Journal of Abnormal and Social Psychology,* 1964, **68**, 248–254.

KUETHE, J. L., & STRICKER, G. Man and woman: Social schemata of males and females. *Psychological Reports,* 1963, **13**, 655–661.

KUETHE, J. L., & WEINGARTNER, N. Male-female schemata of homosexual and non-homosexual penitentiary inmates. *Journal of Personality,* 1964, **32**, 23–31.

PIAGET, J. *Play, dreams and imitation in childhood.* New York: Norton, 1951.

SCHACHTER, S. *The psychology of affiliation.* Stanford: Stanford University Press, 1959.

WEINSTEIN, L. Social schemata of emotionally disturbed boys. *Journal of Abnormal Psychology,* 1965, **70**, 457–461.

(Received October 20, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 435-441

WHEN DOES A HARM-DOER COMPENSATE A VICTIM?

ELLEN BERSCHEID AND ELAINE WALSTER[1]
University of Minnesota *University of Rochester*

It was hypothesized that individuals who had harmed another person would be more likely to compensate their victim if the available compensation made up exactly for the harm done than if, in order to compensate at all, it was necessary to give the victim either an insufficient or an excessive compensation. Ss were members of women's church auxiliaries who in the course of a game were led to deprive a fellow church member of green-stamp books. In a 2nd game they had a chance to award a bonus to the deprived woman. It was found that individuals were more likely to compensate their victim with the bonus if it was adequate to cover the harm done than if the available bonus was insufficient or excessive. From these findings an interesting hypothesis can be derived. If a harm-doer's range of compensatory opportunities is limited, exaggeration by the victim of the harm he has suffered may cause the harm-doer to perceive that his available compensations are inadequate; thus the harm-doer may be less likely to compensate the victim than he would have been had the victim described his suffering in a more modest way.

Common responses to the act of harming another person appear to fall into two major categories: *justification* of the performance of the harmful act and *compensation* to the victim for the harm done.

The distress an individual is likely to feel after harming another may be reduced through the utilization of justification techniques. The harm-doer may, for example, minimize the amount of harm he has done (Brock & Buss, 1962), he may deny responsibility for the harmful act, or he may convince himself that the victim was deserving of his suffering. This last mode of justification as a response to the performance of a harmful act has often been demonstrated by experimenters working within the framework of cognitive-dissonance theory. Davidson (1964), Davis and Jones (1960), and Glass (1964), for example, performed experiments which indicate that the dissonance aroused by performance of a harmful act will be reduced by denigration of the victim, if other means of dissonance reduction are not available.[2] One mode of dissonance reduction which was carefully blocked in all the preceeding studies was compensation. In all cases, the harm-doer's opportunities for atoning to the victim were either nonexistent or minimal.

The possibility for compensation is eliminated from experimental designs so consistently that experimenters sometimes forget that this reaction to harming a victim exists (Aronson, 1967). There are only a few isolated studies which demonstrate that harm-doers do often attempt to compensate the people they have harmed. One such study was performed by Walster and Prestholdt (1966), who examined the effect of public commitment on one's tendency to either justify or compensate for the harm he had done. They found that when the harm-doer was publicly committed to his harmful act, he justified his harm-doing by denigrating the victim. When the subject was not as committed, he tended to compensate the victim by a subsequent overevaluation.

It is obvious that in order for a harm-doer to compensate his victim, opportunities for making compensatory acts must exist. But it also is apparent from daily experience that people do not always choose to avail themselves of their opportunities to make compensation. The experiment reported in this paper was designed to investigate one possible determinant of whether or not a particular compensatory act will be performed. Specifically, we wished to investigate the effect that adequacy of compensation has on the probability of performance of the compensatory act. By adequacy of compensation, we mean the extent to which the compensation benefits the victim, relative to the harm done him.

---

[1] Special thanks are due to Vickey Martins, who was our experimenter, and to Nila Bender, Shirley Damrosch, and Jenny Rajput, who played the role of the confederate. This study was supported in part by National Science Foundation Grant GS 1056 to the authors. We also appreciate the help of Bill Walster and Raymond O. Collier in analyzing our data.

[2] For an explication of this point, see Walster and Berscheid, 1966.

When only one act of compensation is available, we felt that the harm-doer would be more likely to perform that one act when the amount of benefit it would confer upon the victim was adequate (exactly matching the amount of harm done) than when it was inadequate or excessive. In effect we wished to change the focal point of the Mosaic law of justice, "An eye for an eye and a tooth for a tooth," from punishment of the harm-doer to compensation of the victim, and to predict that a compensatory act is most likely to be performed if it will replace an eye for an eye. Acts which will replace a glass eye for one eye, or three eyes for one eye, should have lower performance probabilities. In situations in which the harm-doer's range of compensatory opportunities are limited to acts of these last two types, we predicted that the victim is most likely to be left with no compensation at all.

There is some experimental evidence which bears on this proposition. Walster, Walster, Abrahams, and Brown (1966) found that if an individual discovered that he had accorded another person either less esteem or more esteem than the other deserved, he would subsequently attempt to make up for his injustice. This experiment tends to suggest that people feel others should get exactly what they deserve, no more and no less.

Because we were interested in investigating the extent to which harm-doers use a principle of justice in situations in which compensation is possible, we wished to eliminate from consideration alternative means by which the same hypothesis could be derived. One alternative way in which the hypothesis could be derived is apparent from a consideration of the various functions of reparation. In ordinary circumstances the performance of an act of compensation generally causes the harm-doer himself to suffer or to sacrifice to some extent. And, the amount of suffering involved in making the compensation often increases as the amount of compensation, or benefit conferred upon the victim, increases. Thus, in usual circumstances, the more the harm-doer compensates exactly, the more his suffering is comparable to that of the victim. And if it is true that self-punishment in itself acts to relieve feelings of guilt on the part of the harm-doer, then guilt feelings may be best alleviated by compensating the victim exactly. Suffering less than the victim has suffered does not entirely relieve guilt feelings; suffering more is superfluous. Though this is an interesting hypothesis in itself, we wished to eliminate it from consideration in this experiment. Consequently, the experiment was designed so that the act of compensation would personally cost the harm-doer nothing to perform.

There was one other variable, often present in reparation situations, which we wished to eliminate from our design. The performance of an act of reparation ordinarily gives the victim information that the harm-doer has tried to make up for his injustice. The release of such information may save the harm-doer loss of esteem from the victim and the public at large, and, in addition, may ward off acts of retaliation. Insufficient acts may indicate that the harm-doer has not tried hard enough to compensate and, consequently, may prove unsatisfactory for this reason. Acts which confer excessive benefit upon the victim, on the other hand, may indicate that he has tried very hard to compensate and is generous to boot. In such a case our prediction with regard to excessive compensation would not be likely to hold. We wished to eliminate this confounding consideration from our design. Both insufficient and oversufficient acts should give the same information that the harm-doer has tried to atone, if it is quite clear to all concerned that the one act—inadequate, sufficient, or excessive—is the *only* act in the harm-doer's repertoire.

METHOD

Overview

To test our hypothesis, all subjects inflicted harm on another person and, in all instances, the amount of harm done to the victim remained constant. Specifically, each subject prevented the victim from winning a prize of two books of S&H green stamps in a cooperative game.

Subjects were then divided into three compensation groups: in an *insufficient* compensation group, subjects were given an opportunity to make a compensation which was, in terms of the amount of benefit it would give the victim, inadequate to amend the amount of harm done him. In an *adequate* compensation group, subjects were given an opportunity to exactly compensate the victim. Finally, in an *excessive* compensation group, the only available compensation was one which would benefit the victim over and above the amount of harm done.

Each of the three compensation groups were subdivided, thus yielding a 3 × 2 design. Half of the subjects in each compensation condition (the experimental subjects) were given an opportunity to compensate the person they had harmed. The remaining subjects in each compensation condition (the control subjects) were given a chance to award the compensation only to a person other than the one they had harmed. This latter group served as a control for the extent to which people who have recently hurt another person tend to be generous to other people in their immediate environment, and for the extent to which each level of compensation was a desirable gift for any partner.

In the experimental groups, we predicted that subjects in the *adequate* condition would compensate the victim more often than would subjects in the *insufficient* and *excessive* conditions. We predicted that subjects in the control condition would show less compensation at all compensation levels than experimental subjects, since compensation would not aid the victim. In addition, control subjects should not show the marked differences between the adequate and the inadequate-excessive conditions that we expected in the experimental groups.

Subjects

Subjects were 75 groups of women (240 women in all) who belonged to various church auxiliary groups in the Minneapolis area. The groups ranged in size from two to five women. These women had volunteered to participate in a psychological experiment (which was previously described to the head of each church organization) in return for a donation to the church's treasury.

Procedure

After the women in a particular group arrived at their church along with a female confederate, the experimenter asked all of the ladies to introduce themselves. Quite often all of the women knew each other. They never knew our confederate (Mrs. Bender), who was introduced as a friend of the church secretary and who had been recruited to fill out the group. The experimenter also mentioned, at this point, that one other lady, unknown to the group, a Mrs. Hartig, was scheduled to take part in the survey but that she would be late. Supposedly, the first part of the survey had been explained to Mrs. Hartig on the phone so that the group would be able to start on time and she would be able to begin participating as soon as she arrived.

The experimenter then provided the following rationale for the experiment:

> We are here today to find out how much members of a large urban area such as Minneapolis and Saint Paul know about the community in which they live. You probably have often heard the comment that people generally know a lot about other communities, but very little about their own. In this survey, we are going to try to find out how true that observation really is. This means that we will be asking you questions about the cities of Minneapolis and Saint Paul and the State of Minnesota. In order to make this survey as pleasant as possible, we have decided to gather our information by having you play two games. In each game you will be playing with a partner. As I noted before, there are only five ladies [or how many showed up] here now, but the sixth lady will be the lady that is coming late, Mrs. Hartig. In order to complete the session on time, however, we will have to go right on without her. As soon as she comes she will be assigned a partner.

The experimenter then described two games to the subjects. Ostensibly, the first game gave subjects a chance to win S&H green stamps for themselves and their partners.[3] The second game gave them a chance to win green stamps for a crippled child. (Each woman had earlier been assigned the name of a particular child for whom she would be winning stamps.)

Actually, the first game was designed solely to lead all subjects to deprive their partner of two books of green stamps. The second game was designed solely to give the subject a chance to compensate either her partner or a stranger for the harm the subject had done. The first game was described as follows:

> In the first game you will be randomly assigned to be the partner of another lady in the group. One of the partners will be the supervisor, and the way we will decide this is by a flip of a coin. The supervisor will make all the decisions in the game. I will give the supervisor a set of ten sample questions. As I mentioned before, these will be about Minneapolis and Saint Paul and about Minnesota. The supervisor must then decide, from looking at these sample questions, how many questions of this type she thinks she and her partner can answer when the real game starts. Taking into

---

[3] The authors would like to express their appreciation to the Sperry and Hutchinson Company and to Eugene Beem for the loan of 40 books of green stamps.

> consideration the number she and her partner can answer, she will decide how many books of green stamps she will be going for.

At this time the experimenter distributed a small 3 × 5 card on which there were listed the number of questions each supervisor could set as her and her partner's "quota" of correctly answered questions, and the number of green stamps they would get as a prize if they fulfilled their quota. The higher the quota the more green stamps women could win as a prize. She explained to them:

> If you think, for example, that both you and your partner can each get from 5 to 7 questions right, then you should set your quota for 6 books of green stamps. This means that *both* you and your partner must, individually, get at least 5 questions right. If you set your quota for 5 to 7 questions and it turns out that you get 8 questions right, but your partner only gets 4 right, neither one of you will win the green stamps. You must each get *at least* the number you say you are going for. If you set your quota for 5 to 7 questions and both of you get 8 right, you will still only get the 6 books of green stamps for which you contracted. You can see that it is important that you pick exactly the number of questions that you think that you can get right and set your quota near that level. Are there any questions up to this point?

Then the experimenter explained that each one of the ladies would be working in a separate room so as to prevent conversation. She randomly assigned each woman to a room. After the ladies were ensconced in their rooms, the experimenter went into each room, flipped a coin, and announced that the flip determined that the subject would be a supervisor and that Mrs. Bender (our confederate) would be her partner. She further instructed the subject as follows:

> There is just one more thing that I would like you to do before you start the game. That is, we would like you to write on this piece of paper something about yourself, your hobbies, your interests, or your family. I will take this to your partner. She is doing the same thing. This procedure will enable you to know your partner better, and her to know you a little better, and you both to feel more at ease playing with each other.

A few minutes later, the experimenter brought the confederate's "Personal Interest Card" which stated:

> Nila Bender. Married, have two sons. Moved to Minneapolis ten years ago. Worked as a secretary downtown for two years. Was glad to quit when I got married. I went to college two years before that. Husband went to the University before we were married. Now works for an insurance company. I wish I could see the set of sample questions to tell you what I know about. I know a little about politics, religion and literature (read quite a bit). Am active in the League of Women Voters. If you don't mind I would like to make a suggestion. Let's go for the 1 to 3 question category. My youngest son's birthday is a week from today and I could get him what he wants with two more books. I never thought I would be able to save enough by then but it looks like maybe I have got a chance now. However, you know more about what to go for since you have seen the sample questions.

After distributing these cards and the sample questions to each of the subjects, the experimenter went back to the rooms and asked the subject what quota she had decided to set. Which quota to set posed a problem for the subject. On the one hand she wanted to set a low quota as her partner had suggested. On the other hand, she was anxious to win as many green stamp books as possible for herself. The sample questions the subject received were designed to lead her to set a very high quota. Of the 10 sample questions, 8 were quite easy to answer. They were concerned with things about Minneapolis-Saint Paul that most people should and do know. Two of them, however, were extremely hard. A subject was thus led to the conclusion that she could probably answer 8 of the 10 test questions correctly. All but four of our subjects did disregard the confederate's suggestions and did set a fairly high quota. The data obtained from the four subjects who set the lowest quota are not included in any analyses.

When the test questions arrived it was evident that this strategy had been a mistake. Through pretesting we had selected 10 questions, the answers to which people *should* know, but in fact often do not know (e.g., "Who is the Congressman from your district?"). When all subjects had finished answering their questions, the experimenter picked up their papers, took time to supposedly score them, and again returned to each room. She commented:

> We have now finished scoring your papers and I can give you the scores. You got a score of *two* which, unfortunately, isn't enough. However, your partner did quite well. She got a score of ——— [the confederate was always said to have correctly answered enough questions to meet whatever quota the subject had set]. So even though your partner did answer enough questions correctly, you failed to answer enough questions right. This means that neither of you will receive any stamps at all.

(Pretest interviews showed that subjects felt quite guilty and upset about this turn of events. The score of 2, which all subjects were told they had attained, was enough to win the two books Mrs. Bender had urged the supervisor to try for. Nearly all pretest subjects mentioned this fact. In addition, many commented that through their own "greed" they had deprived their partner of the two books which she very much needed to get a toy for her child's birthday.)

The experimenter then introduced the second game. If the subject had been assigned to the experimental condition, she was told that during the second game she would have the same partner she had had during the first game. If the subject was in the control condition, the experimenter explained that she would have a different partner than she had had in the first game; rather than working again with Mrs. Bender, she would be working with Mrs. Hartig, the lady who supposedly arrived late.

In this game, the experimenter explained that this time the supervisor would not answer any questions concerning Minnesota. Instead, Mrs. Bender (or Mrs. Hartig) would answer the questions and the subject would decide whether her answer was correct or not. She was to put a "C" by the answers that seemed right and a check by any that seemed wrong. Since each question counted 1 point, her paper could receive a grade from 1 to 10.

The experimenter then gave the subject the information that she could compensate her partner if she wished. After reminding the subject that in this game she might have the opportunity to award some green stamps to the crippled boy who had been assigned to her (Johnny Trogan),[4] the experimenter said, as though it was an afterthought:

> One of the ladies felt that she should be able to give the green stamps to her partner instead of Johnny. I'd never thought of that before, but since you are the supervisor you can do whatever you want. It is completely up to you to decide whether to give the stamps to Mrs. __________ [partner's name] or to Johnny . . . that is if you win any stamps to award.

Whether or not the subject was able to award any stamps appeared to be determined by chance. The subject was told to draw a slip from a grab box to see if she would be allowed to award stamps. The woman drew. In each case she drew a "yes," indicating she would have a chance to award stamps. How many stamps she could award was also said to be determined by chance. The experimenter pointed to a second grab box which was said to contain slips of paper indicating the amount of the bonus she could award. Possible bonuses were said to range from 1 stamp to 10 books. Actually only slips containing "three stamps," "two books," or "five books" were in the bag. Once again the subject drew. Depending on which slip they drew, subjects were assigned to the insufficient, adequate, or excessive condition, respectively.

After a short delay, the experimenter returned with the partner's answers to the 10 questions, and the subject was asked to score them. Finally, the experimenter suggested that the subject write down on an address label the name of the person she wanted to receive the green stamps—her partner or the crippled boy. Subjects were reminded that their partners would never know whether or not they had drawn a bonus slip. During the preexperimental discussion, it also had been made clear that Mrs. Bender (and Mrs. Hartig) were not from their neighborhood and so they would not be likely to have a chance to compensate her privately.

Finally, after indicating who should receive the stamps, subjects were asked to fill out a questionnaire which was designed to tap several modes the women might use to justify the harm they had done. The experimenter explained that everyone who was participating in the survey would fill out the questionnaire, designed to tell something about them and their partner, so that we could improve the survey next time. Subjects were instructed to put neither their name nor their partner's name on the evaluation sheet and to put all of the papers in an envelope, which was addressed directly to the research laboratory, where it was to be sent without being opened. Each subject was asked to indicate how much she liked her partner, how much the partner liked her in return, how friendly the partner was, and whether or not she would like to meet her partner. Subjects were also asked to estimate how much the partner appreciated green stamps, how much the boy appreciated green stamps, how much she thought the partner had enjoyed the study, and how much she had enjoyed herself. The degree to which each subject justified her act on these measures were combined to form her total justification score.[5]

After the subject had filled out the questionnaire, the experiment was complete. The experimenter introduced the confederate and explained the deception and the necessity for it.

RESULTS AND DISCUSSION

It will be recalled that we predicted that experimental subjects would be less likely to compensate their victim when the available compensation was insufficient or excessive than when it was adequate.[6] Control groups were run to control for an individual's tendency to compensate anyone—victim or not—

---

[4] As the reader will probably guess, Johnny is simply a device designed to make it possible for the subject to decide not to compensate her partner. If the subject had simply been asked to "Compensate her partner or not," choosing not to compensate would probably have been an unacceptable response to all subjects.

[5] Obviously, interpretation of these justification scores would be difficult since the justification information was always collected after the dependent variable. We asked for this information only as a pretest for a subsequent experiment, designed to test directly the relationship between compensation and justification responses.

[6] When considering the compensation data, the group, not the individual, will be our unit of analysis. Since all group members reported at the same time and were run in the same condition, the average group response is considered as a single observation.

TABLE 1

PERCENTAGE OF SUBJECTS CHOOSING TO COMPENSATE IN VARIOUS CONDITIONS

Object of compensation	Adequacy of possible compensation	(*N*)	% of *S*s compensating
Victim	Insufficient	(13)	42%
	Adequate	(13)	73%
	Excessive	(11)	61%
Nonvictim	Insufficient	(13)	2%
	Adequate	(12)	12%
	Excessive	(13)	26%

in the game situation under different compensation levels. Thus, we expected a larger difference in tendency to compensate between the experimental group and the control group in the adequate condition than in the insufficient or excessive conditions. The appropriate test for our prediction was a quadratic interaction contrast:

$$\Gamma_1 = [+1(\mu_{\text{Exp-Insuf}}) - 2(\mu_{\text{Exp-Adq}}) + 1(\mu_{\text{Exp-Exc}}) - 1(\mu_{\text{Cont-Insuf}}) + 2(\mu_{\text{Cont-Adq}}) - 1(\mu_{\text{Cont-Exc}})]. \quad [1]$$

We tested the hypothesis H_o: $\Gamma_1 = 0$ against the alternative H_a: $\Gamma_1 \neq 0$. An explanation of this procedure is available in Hays (1963). When we turn to our data (See Table 1) we see that our prediction is supported.

In the experimental groups, subjects in the adequate condition are more likely to compensate the victim than are subjects in the insufficient or excessive conditions. In the control conditions this tendency does not exist. In all control conditions, subjects are less likely to compensate their partners than are experimental subjects ($F = 67.36$, $df = 1$, 69, $p < .001$). The more stamps the subject has to award, the more she tends to give them to her partner. The quadratic interaction contrast (Γ_1) we predicted is significant ($F = 3.83$, $df = 1$, 69, $p = .05$).

We also ran a test to determine whether the subjects in the insufficient condition were more or less likely to compensate their partners than subjects in the excessive condition when compared to their relevant control groups. The appropriate test to detect such a linear interaction difference was again a contrast:

$$\Gamma_2 = [+1(\mu_{\text{Exp-Insuf}})0(\mu_{\text{Exp-Adq}}) - (\mu_{\text{Exp-Exc}}) - 1(\mu_{\text{Con-Insuf}}) 0(\mu_{\text{Con-Adq}}) + 1(\mu_{\text{Con-Exc}})]. \quad [2]$$

From this *F* it is clear that individuals are just as unlikely to compensate their partner when the only compensation is insufficient as when it is excessive ($F = .12$, $df = 1$, 69, $p = .73$).

It appears that in every respect our hypothesis is supported.[7] Adequacy of compensation makes an individual especially willing to compensate his victim, and inadequate or excessive compensations appear to be equally discouraging to one's tendency to compensate.

These results have an interesting implication. Both in law suits and in more informal interpersonal relations, exaggeration of the degree of harm one has suffered seems to be a common strategy designed to increase the amount of compensation one will receive from the harm-doer. Such a strategy would be effective if the harm-doer's repertoire of compensatory acts includes one which would exactly compensate the victim for the magnified level of harm. If, however, the harm is exaggerated to such an extent that its level exceeds the highest level of compensation available to the harm-doer to perform, the victim may increase his probability of getting nothing at all. Thus, an errant husband may not attempt any compensation at all to a person whom he has "robbed of the best years of her life," but may compensate if the act in question were said to result in harm of a lesser degree, say just enough harm to be compensated for by a new dishwasher.

---

[7] Since in the control groups the percentage of subjects compensating is very small (2%–26%), while in the experimental groups the percentage is relatively large (42%–61%), the reviewers correctly suggested that an arc-sine transformation of the percentages is appropriate. When such a transformation is made, the results we report remain essentially the same: $p\ [\Gamma_1 = 0 \mid \text{data}] = .07$ ($F = 3.46$, $df = 1$, 69), and $p\ [\Gamma_2 = 0 \mid \text{data}] = .72$, *ns* ($F = .13$).

REFERENCES

ARONSON, E. Dissonance theory: Progress and problems. In T. Newcomb, R. Abelson, W. McGuire, M. Rosenberg, & P. Tannenbaum (Eds.), *Source-*

book on cognitive consistency. New York: Rand McNally, in press.

BROCK, T. C., & BUSS, A. H. Dissonance, aggression, and evaluation of pain. *Journal of Abnormal and Social Psychology,* 1962, **65**, 192–202.

DAVIDSON, J. Cognitive familiarity and dissonance reduction. In L. Festinger (Ed.), *Conflict, decision, and dissonance.* Stanford: Stanford University Press, 1964.

DAVIS, K. E., & JONES, E. E. Changes in interpersonal perception as a means of reducing cognitive dissonance. *Journal of Abnormal and Social Psychology,* 1960, **61**, 402–410.

GLASS, D. C. Changes in liking as a means of reducing cognitive discrepancies between self-esteem and aggression. *Journal of Personality,* 1964, **32**, 520–549.

HAYS, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963. Pp. 462–463.

WALSTER, E., & BERSCHEID, E. Compensation or justification?: The observer's reaction to the victim. Paper presented at the meeting of the American Psychological Association, New York, September, 1966.

WALSTER, E., & PRESTHOLDT, P. The effect of misjudging another: Overcompensation or dissonance reduction? *Journal of Experimental Social Psychology,* 1966, **2**, 85–97.

WALSTER, E., WALSTER, B., ABRAHAMS, D., & BROWN, Z. The effect on liking of underrating or overrating another. *Journal of Experimental Social Psychology,* 1966, **2**, 70–84.

(Received October 24, 1966)

Manuscripts Accepted for Publication in the *Journal of Personality and Social Psychology*

Perceptual Maps of the World: The View from Academia versus the Man in the Street: John P. Robinson and Robert A. Hefner*: Department of Psychology, The University of Michigan, Ann Arbor, Michigan 48104.

The Modification of Dominant Behavior in School Children: Elliot R. Blum and Wallace A. Kennedy*: Department of Psychology, The Florida State University, Tallahassee, Florida 32306.

The Elusive Guilt Reaction: Another Negative Finding: I. W. Silverman*: Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada.

Stability of Cognitive Style from Childhood to Young Adulthood: Herman A. Witkin,* Donald R. Goodenough, and Stephen A. Karp: Psychology Laboratory, Downstate Medical Center, 450 Clarkson Avenue, Brooklyn, New York 11203.

Imitation in Children as a Function of Perceived Similarity to a Social Model and Vicarious Reinforcement: Mary A. Rosekrans,* Institute of Child Development, University of Minnesota, Minneapolis, Minnesota 55455.

Frequency of Contact and Stereotyping: Harry C. Triandis* and Vasso Vassiliou: Department of Psychology, University of Illinois, Urbana, Illinois.

Attitude and Selective Learning: Where are the Phenomena of Yesteryear? Anthony G. Greenwald* and Joseph S. Sakumura: Department of Psychology, The Ohio State University, 1945 North High Street, Columbus, Ohio 43210.

Validation of Stereotypes: Japanese versus American Students: F. K. Berrien*: Department of Psychology, Rutgers University, New Brunswick, New Jersey 08903.

---

* Asterisk indicates author for whom address is supplied.

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 442-447

BRIEF ARTICLES

COMMITMENT TO SELF-APPRAISAL AND INTERPERSONAL EVALUATIONS

STEPHEN C. JONES AND CARL RATNER

State University of New York at Buffalo

How does commitment to a low self-appraisal affect interpersonal evaluations? In an experiment designed to examine this question Ss were induced by false test scores to adopt low ability appraisals, then were either given the opportunity to commit themselves to this appraisal or were given no such opportunity. Subsequently Ss participated in a 3-person social exchange situation in which they were positively evaluated by 1 peer and negatively evaluated by another peer. The results showed that Ss given the opportunity for commitment more positively evaluated a peer sending positive evaluations than a peer sending negative evaluations. Ss not given the opportunity for commitment more positively evaluated a peer sending negative evaluations than a peer sending positive evaluations. The results were interpreted as indicating that commitment to a low ability level may protect a person from the behavioral implications of accepting praise and rejecting censure.

A frequently observed tendency in everyday social behavior as well as laboratory investigations of social interaction (Byrne, 1961; Dittes, 1959; Jones, 1966) is for a person to positively evaluate another person who approves of him or agrees with him and to negatively evaluate another person who disapproves of, or disagrees with him. Some studies have shown, however, that this general tendency toward reciprocity can, under certain conditions, be eliminated or even reversed when a person has a low appraisal of himself (Deutsch & Solomon, 1959; Gerard, 1961, 1965; Gerard & Rotter, 1961; Jones & Schneider, 1966; Wilson, 1965). This contrary finding, that a self-derogator tends to respond more favorably to social censure than to social praise, has been interpreted as supporting the consistency theories of Lecky (1945), Festinger (1957), Heider (1958), and Secord and Backman (1961). Praise is inconsistent with a low self-evaluation and is reacted to negatively, and censure is consistent with a low self-evaluation and is reacted to positively. The general problem is to identify those variables that distinguish between consistent interpersonal responses and reciprocal interpersonal responses. One such variable may be the extent to which a person is committed to his self-appraisal.

The present study explored the effects of commitment to a low self-appraisal on a person's interpersonal reactions under conditions in which he received positive evaluations from one peer and negative evaluations from another peer. Two conceptions of the role of commitment were considered, one suggested by consistency theorists and one which arose from our own investigations.

In discussing the relationship of commitment to consistency theories, and especially cognitive dissonance theory, Brehm and Cohen (1962) argued that commitment increases resistance to change of a cognitive element and that "dissonance aroused in connection with a commitment is likely to be reduced by change in elements other than those involved in the commitment [p. 9]." Applying this conception to the present paradigm, one would expect that the extent of a person's commitment to a low self-appraisal would be directly related to the tendency to devaluate a peer sending positive evaluations relative to a peer sending negative evaluations.

Our own work has led to another view of commitment. When exchanging evaluations with others, people may well be concerned with the consequent behaviors implied in the kind of evaluative information they accept for themselves. Accepting praise from others implies that one expects to participate effectively in difficult activities; accepting censure from others implies that one expects to participate ineffectively in difficult activities. We propose that the more a person is committed to a low self-appraisal, the less important are these consequences for his evaluative reactions to others. If a person has made no commitment to a low self-appraisal, then accepting praise or rejecting censure from another person implies choosing a level of performance beyond his capabilities with the hazard of experiencing failure. However, if a person is committed to a low self-appraisal by the decisions he makes or by public acknowledgment, then he is protected from these negative implications of accepting praise. In effect, he has avoided the choice of difficult situations implied in accepting the esteem of others by irrevocably committing himself to easy situations which are commensurate with his low self-evaluation. In this case the person can accept the esteem without simultaneously risking the undesirable consequences of "living up to" its implications. If commitment has such a protective function, then one would expect that, contrary to the consistency analysis, the extent of a person's commitment to a low self-appraisal would be inversely related to the tendency to devaluate a peer sending positive evaluations relative to a peer sending negative evaluations.

METHOD

Subjects

Forty-six female subjects who volunteered from introductory psychology classes as part of the course requirement were run in groups of three in experimental sessions lasting approximately 1 hour. Occasionally a subject did not appear, and a graduate student was paid to complete the three-person group; her data were not included in the analyses.

Apparatus

Subjects were seated in separate cubicles which visually isolated them from one another. In front of every subject was a panel with two columns, each containing one green light and one red light for receiving positive and negative evaluations from peers and one double-throw switch for sending positive and negative evaluations to peers. A master control panel allowed the operator[1] to record from Mercury counters the number of positive and negative evaluations made by each subject and to feed back to subjects by means of double-throw switches a prearranged schedule of green and red lights.

Procedure

The experiment was presented as an analysis of the predictive validity of a recently developed measure of "personality inference" ability. Subjects were informed that during the first experimental session they would be taking written and oral parts of the Feldman-Collier Personality Inference test and that during a second experimental session they would be asked to analyze the characteristics of a mental patient as revealed in a tape-recorded clinical interview. Their analysis was to be evaluated by a clinical psychologist who would be present at the session, and the relations between the clinician's evaluation and their test scores would be studied.

Subjects then took four sections of the written part of the Feldman-Collier test, which is actually fictitious. Each item consisted of two value statements containing the answers purportedly given by a particular person and one unanswered value statement, as illustrated in the sample item:

1. undecided: Avoid depending upon other persons and things.
2. true : Truth existed before man.
3. ________ : It's a wonderful feeling to sit surrounded by your possessions.

The subject's job was to draw inferences about the person presumably represented in the item on the basis of the answered statements, then to predict how that same person responded to the unanswered statement. These predicted answers could be "true," "undecided," or "false."

After each of the four written sections of the Feldman-Collier test, subjects received scores indicating how well they had done. So that they would adopt a low self-appraisal, subjects in both the commitment and no commitment conditions received scores on these written sections of 33%, 50%, 17%, and 33% correct, respectively.

Manipulation of commitment. After taking the written part of the test, subjects were informed that for the second experimental session there was a series of tape-recorded clinical interviews which differed in complexity and difficulty. In the no-commitment condition subjects ($N = 20$) were told that *at the conclusion* of the oral part of the test they would be able to choose which one of the clinical cases they wished to analyze during the second session. In the commitment condition subjects ($N = 20$) were given a list of the clinical cases, ranked in order of difficulty, and were asked to choose *immediately* which case they wished to analyze during the second session. The experimenter mentioned that it was necessary to make the choice before the oral part of the test so that schedules could be arranged with the clinical psychologist and subjects could be informed of the schedule prior to leaving the first experimental session. The experimenter also stated that the choice of clinical case would be irrevocable.

Control condition. In order to check the low self-appraisal induction and the assumed relation between self-appraisal and choice of task, six subjects (two groups) were given written test scores of 83%, 100%, 67%, and 83% correct and were run in a control condition which in all other respects was similar to the commitment condition.

Manipulation of evaluations received from peers. During the group session, which constituted the oral part of the test and followed the written part, the manipulation of the evaluations received from others was introduced, and the main dependent variable, the evaluations sent to others, was measured. The items in the oral part were from the Feldman-Collier test, but were not the same items used in the written part. The subjects were informed that

---

[1] Our thanks to Ellen Abelson for her assistance in operating the master panel.

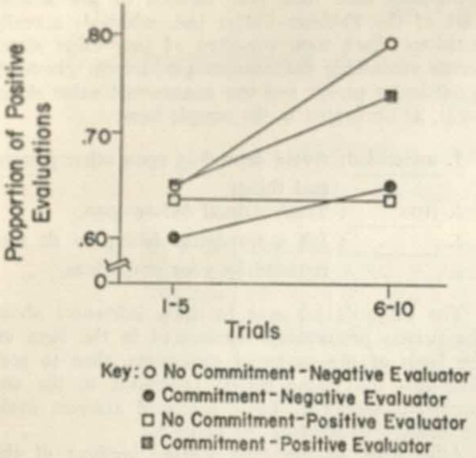


FIG. 1. Evaluations sent to others as a function of commitment, evaluations received from others, and blocks of five trials.

the evaluations received in this session would come not from an expert source, but from the other group members and that the rationale for this procedure was "that psychologists have recognized that a person's social sensitivity may depend to some extent upon the kind of people with whom she is associating." In this group session there were 10 trials; on each trial each subject gave aloud one test answer, and the other two subjects simultaneously evaluated the answer either positively or negatively with the double-throw switches on their panels. A positive evaluation signified that the answer given was thought to be correct, and a negative evaluation signified that the answer given was thought to be incorrect. The red and green light evaluations that a subject received on her panel were experimentally controlled so that each subject received 9 out of 10 green lights purportedly from one peer (positive evaluator) and 4 out of 10 green lights purportedly from the other peer (negative evaluator). The ordering of red and green lights over trials and the ordering of positive and negative evaluators relative to the responding position of the subject were varied to reduce systematic order biases.

The evaluations subjects sent to one another were intercepted on the master control panel and recorded by the operator. These switch-throwing evaluations were the main dependent variables of the study.

TABLE 1

PROPORTION OF POSITIVE EVALUATIONS SENT TO OTHERS AS A FUNCTION OF COMMITMENT AND EVALUATIONS RECEIVED FROM OTHERS

	Positive evaluator	Negative evaluator
Commitment ($N = 20$)	68.9	61.9
No commitment ($N = 20$)	64.1	71.6

Further data obtained were as follows: (*a*) After the written part of the test, subjects were administered a questionnaire to assess their own ability ratings; and (*b*) after the group session, subjects were administered questionnaires to again assess their own ability ratings and also their perceptions of the other group members.

At the conclusion of the experimental session the manipulations were disclosed, the subjects were informed that there would be no further session, and the real purpose of the experiment was discussed.

RESULTS

Establishment of Appropriate Conditions

In order for the theoretical analyses discussed earlier to be applicable to the experimental situation it was necessary to show, first, that subjects had adopted low self-appraisals and, second, that those subjects in the commitment condition had not only chosen from among the easiest clinical tasks but also that these choices were the result of subjects' low self-appraisals. Evidence that the first condition was met is the fact that the initial self-ratings of subjects in the commitment and no-commitment conditions averaged 2.55 on a 7-point scale (on which 1 was low and 7 was high) as compared to an average self-rating of 4.75 made by subjects receiving high written test scores in the control condition ($U = 12.5$; $p < .01$). The final self-ratings for subjects in the commitment and no-commitment conditions were somewhat higher though not significantly different from the initial ratings. None of the self-rating means was significantly different from any other as a function of the commitment and no commitment conditions.

Supporting the second condition stated above is the finding that in the commitment condition 18 subjects chose from among the least difficult clinical cases, 2 chose moderately difficult cases, and none chose from among the most difficult cases, whereas in the control condition no subject chose from among the least difficult cases, 5 subjects chose moderately difficult cases, and 1 subject chose from among the most difficult cases. The most difficult and moderately difficult cases were combined into one category, and the difference between the distributions of the commitment and control conditions was found to be significant ($p < .001$; Fisher exact test). This result lends support to the assumption that by choosing the easiest task a subject was implicitly committing herself to a low self-appraisal.

Analysis of Switch-Throwing Evaluations

For each subject the proportion of positive evaluations to total evaluations sent to the positive evaluator and negative evaluator, respectively, was computed. The means for these data are presented in Table 1, and the same data divided into two blocks of five trials each are presented in Figure 1. Subjects in the commitment condition more positively evaluated the positive evaluator than the negative evaluator, and subjects in the no-commitment condition more positively evaluated the negative evaluator than the positive evaluator. Considering the entire 10 trials this interaction is significant at the .05 level ($F = 4.07$; $df = 1/38$), and for the second block of five trials the interaction is significant at the .01 level ($F = 8.45$; $df = 1/38$).

Analysis of Questionnaire Evaluations

At the conclusion of the experiment subjects were asked to rate both peers with regard to liking, respect, candidness, competence, and effectiveness as a rater of personality-inference ability. The mean responses to these questions appear in Table 2. Although the pattern of mean questionnaire responses as a function of commitment and evaluations received from others was in a direction similar to the switch-throwing evaluations, neither the interactions nor the main effects was significant.

TABLE 2

MEAN QUESTIONNAIRE EVALUATIONS OF PEERS

Dimension	Commitment		No commitment	
	Positive evaluator	Negative evaluator	Positive evaluator	Negative evaluator
Liking	5.15[a]	4.12	5.15	4.65
Respect	5.05	4.06	5.15	4.80
Candidness	5.60	4.94	5.30	5.40
Competence	4.25	3.88	4.35	4.55
Skill as rater	4.15	4.00	4.45	4.30

[a] For all means the higher the number the more favorable the rating.

DISCUSSION

These results may be interpreted as support for the notion that commitment to a low self-appraisal may protect a person from the undesirable implications of accepting social praise and rejecting social censure. People appear to like praise, and they will respond favorably to it, provided that in so doing they are not also accepting its contingent responsibilities. Commitment to a low self-appraisal can effectively eliminate these responsibilities, and under such conditions praise can be accepted with impunity.

The strength of this interpretation depends to some extent on what the switch-throwing responses meant to the subjects. Two inferences about that meaning could lead to conclusions which support the Brehm and Cohen (1962) analysis of commitment. First, suppose the subject assumed she was being scored on the correctness of her switch-throwing responses. Since the subject presumably viewed her own test answers as incorrect, she might have concluded that the positive evaluator was also giving incorrect test answers but that the negative evaluator was giving correct test answers. By positively evaluating the positive evaluator relative to the negative evaluator, the subject was intentionally making incorrect evaluations and thereby being consistent with her low self-appraisal. Hence, subjects in the commitment condition were actually behaving more consistently than subjects in the no-commitment condition. This analysis rests on the assumption that the negative evaluator was perceived as more competent than the positive evaluator, and the fact that this assumption was not supported by the data (see Question 4 in Table 2) indicates that this interpretation is not tenable.

A second possibility is that subjects were using the switch-throwing responses to ingratiate or "win over" the other group members. Paradoxically, the person toward whom subjects were most favorable would be the person whose opinion they most wanted to change. Hence subjects in the commitment condition were behaving more consistently than subjects in the no-commitment condition by showing less desire to change the negative evaluator's opinion. Such tactics were doomed to failure, however, by virtue of the experimental control of the evaluations received, and therefore subjects would be expected to show a more negative reaction toward peers who had not been won over. However, subjects' questionnaire evaluations of others, especially on the dimensions of liking and respect, were in a direction opposite to this expectation (see Table 2). These questionnaire evaluations could not be tactical since it was clear to the subjects that they would not be received by peers; therefore, the second interpretation is not supported by the data.[2]

We are inclined to the view that these switch-throwing responses were generally affective in nature and indicated the extent to which the subject was made uncomfortable by the evaluations she received from each peer. The more uncomfortable the subject felt as a result of the evaluations she received, the less the proportion of positive evaluations she sent. Furthermore, receiving positive evaluations was pleasant unless, as in the no-commitment condition, it implied the consequence of choosing a more difficult clinical task.

---

[2] This interpretation arose from an idea presented by Elliot Aronson in his recent grant proposal concerning, "Some Causes and Effects of Esteem and Rejection," unpublished manuscript, University of Texas, February 1966.

This interpretation leads to the conclusion that the data of the present experiment are opposite to what one would expect from Brehm and Cohen's (1962) analysis of the role of commitment in cognitive dissonance theory and also Festinger's (1964) analysis of postdecision dissonance. However, the weight of evidence supporting both of these cognitive consistency analyses (see, e.g., Brehm & Leventhal, 1962; Brock, 1965; Cohen, Greenbaum, & Mansson, 1963; Festinger, 1964; Gerard, 1964; Kiesler & Corbin, 1965; Mills & Ross, 1964) indicates that the notion of commitment as a protective mechanism may be the exception to a more general rule on the functioning of commitment. The problem, therefore, is to specify the conditions under which commitment operates to protect the individual. Characteristics of the present experiment that might have led to this difference in the role of commitment were, first, that the emphasis of the experimental situation was on the self-evaluation of an ability, second, that the subjects were led to believe they were poor on this ability and could consequently fail on tasks requiring the use of the ability, and third, that future events would provide the opportunity for subjects to experience this failure. Under these circumstances subjects may well have been more highly motivated by the desire to gain successes or avoid failures than by the need to be consistent.

The main question to be raised regarding predictions about the functioning of commitment in interpersonal evaluating behavior is whether or not there are specifiable negative events from which the subject can in fact be protected. If the existence or nonexistence of negative events is a crucial mediating variable, then both the consistency notion and the protective notion of commitment might be demonstrated in an experimental paradigm similar to the one used in the present study. Whether or not subjects anticipate the possibility of negative events (i.e., failure on a task) would be manipulated as follows: in one condition subjects must choose a clinical task which they will be expected to perform, and in another condition subjects must choose a hypothetical clinical task which they will not be expected to perform. In the first condition, in which commitment can protect the subject against choosing and performing an overly difficult task, the results of the present study should be replicated, and in the second condition, in which commitment can serve no such protective function, the results predicted from consistency theory would be expected.

REFERENCES

BREHM, J. W., & COHEN, A. R. *Explorations in cognitive dissonance.* New York: Wiley, 1962.

BREHM, J. W., & LEVENTHAL, G. S. An experiment on the effect of commitment. In J. W. Brehm, & A. R. Cohen, *Explorations in cognitive dissonance.* New York: Wiley, 1962. Pp. 192–201.

BROCK, T. C. Commitment to exposure as a determinant of information receptivity. *Journal of Personality and Social Psychology,* 1965, 2, 10–19.

BYRNE, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology,* 1961, **62**, 713–715.

COHEN, A. R., GREENBAUM, C. W., & MANSSON, H. H. Commitment to social deprivation and verbal conditioning. *Journal of Abnormal and Social Psychology,* 1963, **67**, 410–421.

DEUTSCH, M., & SOLOMON, L. Reactions to evaluations by others as influenced by self evaluations. *Sociometry,* 1959, 22, 93–112.

DITTES, J. E. Attractiveness of group as a function of self-esteem and acceptance by group. *Journal of Abnormal and Social Psychology,* 1959, **59**, 77–82.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L. *Conflict, decision, and dissonance.* Stanford: Stanford University Press, 1964.

GERARD, H. B. Disagreement with others, their credibility, and experienced stress. *Journal of Abnormal and Social Psychology,* 1961, 62, 559–564.

GERARD, H. B. Conformity and commitment to the group. *Journal of Abnormal and Social Psychology,* 1964, **68**, 209–211.

GERARD, H. B. Deviation, conformity, and commitment. In I. D. Steiner & M. Fishbein (Eds.), *Current studies in social psychology.* New York: Holt, Rinehart & Winston, 1965. Pp. 263–277.

GERARD, H. B., & ROTTER, G. S. Time perspective, consistency of attitude, and social influence. *Journal of Abnormal and Social Psychology,* 1961, **62**, 565–572.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

JONES, S. C. Some determinants of interpersonal evaluating behavior. *Journal of Personality and Social Psychology,* 1966, 3, 397–403.

JONES, S. C., & SCHNEIDER, D. J. Certainty of self-evaluation and reactions to evaluations from others. Paper read at Midwestern Psychological Association, Chicago, May 1966.

KIESLER, C. A., & CORBIN, L. H. Commitment, attraction, and conformity. *Journal of Personality and Social Psychology,* 1965, 2, 890–895.

LECKY, P. *Self-consistency, a theory of personality.* New York: Island Press, 1945.

MILLS, J., & ROSS, A. Effects of commitment and certainty upon interest in supporting information.

Journal of Abnormal and Social Psychology, 1964, 68, 552–555.

SECORD, P. F., & BACKMAN, C. W. Personality theory and the problem of stability and change in individual behavior: An interpersonal approach. *Psychological Review,* 1961, 68, 21–32.

WILSON, D. T. Ability evaluation, postdecision dissonance, and co-worker attractiveness. *Journal of Personality and Social Psychology,* 1965, 1, 486–489.

(Received July 7, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 447–450

ANTI-SEMITISM, STRESS, AND ANCHOR EFFECTS ON INTERPERSONAL JUDGMENTS[1]

DONALD FISCHER AND BRENDAN G. RULE

University of Alberta

Personality and situational factors related to evaluations of neutral bystanders following stress were investigated by varying prejudice, stress, anchor stimuli, and persons evaluated. Data indicated the following results: (a) Unfavorable evaluations were greater following a negative-anchor than following a positive-anchor condition; (b) more unfavorable evaluations were expressed toward the 1st person evaluated than toward the 2nd; (c) high- and low-prejudiced individuals were more favorable than were moderates towards others under mild stress, whereas the moderately prejudiced individuals were more favorable than the extremists under severe stress; (d) prejudice, stress, anchor, and the persons rated significantly interacted. Results supported the notion that the judgments of the high- and low-prejudiced persons are similar, while the moderately prejudiced individual differs from extremists in judgments. Furthermore, for the 1st person rated, an assimilation-contrast interpretation of the data partially confirmed Berkowitz' theoretical notions about judgmental effects in personality functioning.

The scapegoat theory of prejudice has received considerable experimental attention in the last decade. One particular focus of interest has been with delineating differences in hostility displacement by high- and low-prejudiced persons. In a series of studies, Berkowitz (1959, 1960, 1961) has reported that prejudiced individuals manifest different judgmental processes under stress than do their less-prejudiced peers. These studies indicated that under stress highly prejudiced individuals exhibit assimilation effects or make gross discriminations among stimuli, whereas low-prejudiced individuals exhibit contrast effects or make finer discriminations among stimuli. With respect to the kinds of judgments made, the highly prejudiced persons typically displayed assimilation effects by displacing hostility onto neutral bystanders. In other words, peers were included into the same category as an annoying experimenter, and hostility aroused by the latter was generalized to the former. In this case, the annoying experimenter served as a negative judgmental anchor.

However, when Berkowitz (1961) attempted to demonstrate the generality of his notions about the judgmental processes of the high- and low-prejudiced persons by introducing a positive anchor, only partial confirmation of his hypotheses was obtained. Since his findings are ambiguous and restrict a judgmental process explanation for the behavior of prejudiced persons, one purpose of the present study was to investigate the relevance of the anchor variable in an interpersonal situation rather than in the nonpersonal situation that characterized Berkowitz' study.

Furthermore, a very basic issue pertaining to displacement phenomena is whether high- and low-prejudiced groups do in fact differ from each other in the nature of their judgments. Recently, Rule (1966) found that judgmental effects for the high- and low-prejudiced individuals did not differ from each other but did differ from the moderately prejudiced persons. The extremists were more negative towards neutral strangers following stress than were the moderately prejudiced persons. The present study was designed also to provide further support for and extend Rule's findings regarding similarity of extremists' behavior.

---

[1] This paper is based in part on a master's thesis submitted to the University of Alberta by Donald Fischer. Portions of this research were reported in a paper read at the 1966 Canadian Psychological Association meeting.

The discrepancy between Rule's and Berkowitz' findings was explained by differences in subject-selection criteria. Rule suggested that her moderately prejudiced group was comparable to Berkowitz' low-prejudiced group. Accounting for the inconsistency in findings in this way, it was possible to elaborate expectations consistent with both Berkowitz' and Rule's theoretical positions. It was expected that following stress high- and low-prejudiced persons would manifest similar judgmental processes, but would differ from moderately prejudiced persons. The extremists were expected to exhibit assimilation effects, and the moderately prejudiced persons were expected to exhibit contrast effects in judgments of neutral bystanders. According to Berkowitz' rationale, assimilation to a negative anchor should result in increased hostility, whereas assimilation to a positive anchor should result in increased favorability. Contrast to a negative anchor should be apparent in increased favorability, whereas contrast to a positive anchor should be apparent in increased hostility.

METHOD

Subjects

Subjects were 96 introductory psychology students who participated to meet course requirements. These subjects were selected from a pool of 639 students who had completed a questionnaire which included the 10-item Anti-Semitism and the 28-item F scales (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). For both scales item scores varied from 5 points for a response indicating strong agreement with an item to 1 point for a response indicating strong disagreement.

Median scores for subjects categorized as high prejudiced, moderately prejudiced, and low prejudiced were Anti-Semitism, 29, F, 92; Anti-Semitism, 21, F, 82; Anti-Semitism, 14, F, 65, respectively.

Procedure

Subjects were run in 24 groups of 4 each. The groups were homogeneous with respect to sex and level of prejudice. Subjects were greeted and briefly introduced to each other. Each subject stated his or her name. No more communication was permitted. While subjects were in view of one another, the experimenter indicated seating arrangements. Subjects were then seated in cubicles which restricted their range of observation.

Subjects were told that there were two parts to the study. One task was to work out problems, the other was to provide information about first impressions of people. At the beginning of the session, subjects were given a rating scale consisting of a list of 13 bipolar adjectives. Each subject was asked to complete an evaluation of the person on his immediate right, with the exception of the person seated in the most extreme right cubicle who evaluated the person seated in the most extreme left cubicle. The experimenter collected these ratings ostensibly for distribution to the person rated, but in fact distributed substituted ratings which were either favorable or unfavorable. It was assumed that receiving an unfavorable evaluation of oneself from a peer would induce a negative anchor and receiving a favorable evaluation would induce a positive anchor. This procedure constituted manipulation of the anchor variable.

Subjects were then asked to solve a list of 13 difficult anagrams within an 18-minute period. The anagrams corresponded to the problems used by Sarason (1961). High- and low-stress induction was accomplished by instructions. High-stress instructions indicated that the performance on the task was directly related to intelligence and that high school students of above-average intelligence (IQ greater than 100) were able to successfully complete the task. Low-stress instructions indicated that the anagrams were difficult, and completion of all the anagrams was unlikely.

Subjects were given two rating scales identical to the one they had previously used and were asked to make successive judgments of the persons on their right and second from their left. It was intended that these ratings would depend only on the characteristics perceived during the subjects' brief introduction at the beginning of the session. Before leaving the experimental room, subjects were informed of the true nature of the experiment and asked to maintain secrecy until the completion of the experiment in about 1 month's time.

RESULTS

To assess any possible differences between experimental groups in ability to solve anagrams, the mean number of anagrams correctly solved was examined. Results from an analysis of variance provided no evidence that high-, medium-, and low-prejudiced subjects differed in the number of correct solutions with 18 minutes time ($F = 2.34$, $df = 2/84$).

Evaluations given by the high-, medium-, and low-anti-Semitic subjects were examined. Data consisted of total scores over the 13 adjective scales for each of the two people described. Low scores indicated a favorable evaluation and high scores, an unfavorable evaluation. Data were analyzed using a $3 \times 2 \times 2 \times 2$ analysis of variance, split-split plot design. There were 3 levels of prejudice, 2 levels of stress, 2 levels of anchor, and 2 persons rated.

Results of the analysis indicated the following relationships. More negative ratings followed a negative anchor than a positive one ($F = 15.95$, $df = 1/84$, $p < .01$). The means were 3.16 and

2.67 for the negative and positive anchors, respectively. This indicates that the anchor manipulation was successful.

More hostility was expressed toward the first person evaluated than toward the second person ($F = 7.86$, $df = 1/85$, $p < .01$). The means were 3.07 and 2.82, respectively.

The high- and low-prejudiced persons were more favorable than the moderately prejudiced persons under mild stress, whereas the moderately prejudiced persons were more favorable than the extremists under severe stress ($F = 5.86$, $df = 2/84$, $p < .01$). Table 1 indicates the means for this interaction.

Prejudice, anchor, and person evaluated significantly interacted ($F = 4.31$, $df = 2/85$, $p < .01$). With respect to the first person evaluated, high- and low-prejudiced individuals increased in unfavorable evaluations as stress increased from mild to severe following a negative anchor and decreased in unfavorable evaluations following a positive anchor; the moderately prejudiced persons decreased in unfavorable evaluations as stress increased following both the positive and negative anchor. Thus, high- and low-prejudiced persons exhibited an assimilation to the anchor effect under stress, while the moderately prejudiced persons exhibited a contrast effect following the negative anchor and an assimilation effect only following the positive anchor.

With respect to the second person rated, as stress increased high- and low-prejudiced persons increased in unfavorable evaluations following a positive anchor, but the high-prejudiced individuals decreased in unfavorable evaluations following a negative anchor, and the low-prejudiced subjects increased in unfavorable ratings following the negative anchor. The moderately prejudiced individuals decreased in negativity following both the positive and negative anchor as stress increased. Table 2 contains the mean evaluation scores for the conditions of this interaction. No other main effects or interactions were significant.

DISCUSSION

Data supported Rule's findings that under stress extremely high- and extremely low-prejudiced persons are similar in their evaluations of other people, whereas the moderately prejudiced persons differ in reactions from the extremists. These findings lend further support to Rokeach's (1960) notion that the conceptual systems of extremists are similar. As previously suggested, it appears that as in Rule's study the high- and moderately prejudiced groups in this study compare to Berkowitz' high- and low-prejudiced groups, respectively. It is evident that a large number of subjects must be sampled to adequately specify the categories of high-, moderate-, and low-prejudiced persons.

TABLE 1

MEAN EVALUATION SCORES FOR THE HIGH-, MODERATE-, AND LOW-PREJUDICED GROUPS FOLLOWING MILD AND SEVERE STRESS

Level of prejudice	Mild stress	Severe stress
Low	2.56	2.87
Moderate	3.36	2.76
High	2.98	2.96

Consistent with expectations, the results also indicated that stress served to broaden the categories of the high- and low-prejudiced persons so that more unfavorable evaluations were generalized to the first person described when a negative anchor was provided and more favorable evaluations were generalized when a positive anchor was provided. While the dependence of the high-prejudiced authoritarian individual on the external environment has long been recognized (Adorno et al., 1950), this study demonstrated a similar dependence on external sources by the low-prejudiced person. These data also demonstrated that consistent judgmental effects operate for positive as well as negative stimuli. The implication for extremists is that, at least for the first person responded to, the effects of stress can be channeled into a favorable or unfavorable direction depending on the anchor stimuli provided in a particular situation.

TABLE 2

MEAN EVALUATION SCORES

Level of prejudice	1st person rated			
	Positive anchor		Negative anchor	
	Mild stress	Severe stress	Mild stress	Severe stress
Low	2.88	2.73	2.53	3.15
Moderate	3.21	2.48	3.62	3.23
High	2.91	2.61	3.25	3.50
	2nd person rated			
Level of prejudice	Positive anchor		Negative anchor	
	Mild stress	Severe stress	Mild stress	Severe stress
Low	2.30	2.79	2.53	2.80
Moderate	2.85	2.30	3.76	3.00
High	2.39	2.55	3.38	3.16

Results regarding the second person evaluated did not form a consistent pattern. It appeared that the occurrence of the first evaluation altered the stimulus situation for the second evaluation. Since more negativity was expressed towards the first person rated, it is clear that hostility was reduced on the second rating. This could be accounted for in terms of catharsis, that is, a reduction in the instigation to hostility or in terms of aggression-anxiety—guilt or anxiety arousal which inhibited the expression of further hostility. Regardless of the mechanism involved in the hostility reduction, the conditions necessary to produce the expected judgmental effects in terms of assimilation and contrast did not appear to be present.

Of particular interest is the reaction of the moderately prejudiced person in terms of the nature of his judgments. The medium-prejudiced subjects reacted to increased stress with increased friendliness regardless of the anchor stimuli or persons evaluated. Stress seemed to establish a set to respond favorably regardless of other situational factors. The moderately prejudiced person appeared to respond more to internal evaluational cues rather than to external situational cues as did the extremists.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BERKOWITZ, L. Anti-Semitism and the displacement of aggression. *Journal of Abnormal and Social Psychology,* 1959, 59, 182–187.

BERKOWITZ, L. Judgmental processes in personality functioning. *Psychological Review,* 1960, 67, 130–142.

BERKOWITZ, L. Anti-Semitism, judgmental processes, and displacement of hostility. *Journal of Abnormal and Social Psychology,* 1961, 62, 210–215.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

RULE, B. G. Anti-Semitism, stress, and judgments of strangers. *Journal of Personality and Social Psychology,* 1966, 3, 132–134.

SARASON, I. G. The effects of anxiety and threat on the solution of a difficult task. *Journal of Abnormal and Social Psychology,* 1961, 62, 165–168.

(Received July 13, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 450–453

ACCURACY OF JUDGMENTS OF PERSONAL ATTRIBUTES FROM SPEECH [1]

RAYMOND G. HUNT AND TIP KAN LIN

State University of New York at Buffalo

By asking judges in several conditions to predict speakers' responses to an adjective check list, data were generated bearing on the accuracy of perceptions of personality from voice. These data are discussed in relation to voice quality as an expression of personality and also in relation to questions of generalized abilities in person perception.

With regard to questions of accuracy in the interpersonal judgment of personality, Gordon Allport put the matter plainly when he stated in his *Pattern and Growth* volume (1961) that:

> Although it is certainly true that the judging process is complex, and that it is hard to find criteria for measuring the success of judgments, the fact remains that accuracy of judgment is a major goal in our social relationships, and that some people make more correct judgments than others [p. 513].

So, despite the well-known limitations of such research, it seemed to be worth an effort to explore further some variables and issues germane to the accuracy of judgments of personality, in this case, from speech.

In particular, the authors wanted to investigate: (*a*) whether judges can judge personality accurately from speech, (*b*) whether judges are consistently accurate in their ratings across samples of speech produced by dissimilar personalities, (*c*) whether there are individual differences in judgmental accuracy and whether personality traits differ in their ease of accurate apprehension, (*d*) whether speakers performing in-character verbal parts will be perceived more accurately than the same speakers heard in an out-of-character part, (*e*) whether open-minded judges perform differently from closed-minded ones, and (*f*) whether interpersonal similarity is related to accuracy of judgment.

---

[1] This paper was adapted from a master's thesis submitted by the second author to the graduate school at the State University of New York at Buffalo. A version was presented at the Eastern Psychological Association, New York, April 1966.

In brief, the study dealt with accuracy of judgment in relation to aspects of the judgment task, of the stimulus person (i.e., the speaker), of the judge (i.e., the listener), and of the relation between the latter two. It was assumed that data would be generated relevant to questions of relations between voice and personality on the one side and the controversial issue of generalized ability in person perception on the other.

METHOD

Subjects

The subjects or judges were 24 female and 54 male undergraduates from a course in social psychology. Most of them were juniors majoring in psychology, and they ranged in age from 19 to 24 years except for one subject who was 39.

Procedure

During a regular class period the judges were given the Dogmatism E Scale. None of the subjects connected administration of the test with their later participation in the experimental phase of the research. For the latter phase, a closed-minded group was composed of 28 subjects scoring one standard deviation or more above the total E scale mean, and an open-minded group was composed of 24 subjects scoring one standard deviation or more below the mean. The remaining 26 subjects formed an intermediate group. Within each group, judges were assigned at random to one of two conditions.

In each of those conditions the judges, in groups of two to eight, listened to two different male speakers reading the same prose passage. However, in Condition I the first speaker performed an in-character part while the second speaker performed out of character. That is, Speaker A read a passage he had previously described to the experimenters as being "like me"; Speaker B read the same passage which he had described as "not like me." In Condition II, using a different passage, these circumstances were reversed, although the order of speakers remained constant.

The passages read by the speakers were two of the three developed by Markel and Roblin (1965) in their study of the effects of content upon judgments of personality from voice. The two passages used differed sharply in content, one being Markel and Roblin's "death-hostility" passage, and the other their "vocational-neutral" one. The two speakers (and, in effect, the two passages) were selected by asking 16 graduate students to review the three Markel-Roblin passages, checking the one "most like me" (in character) and the one "least like me" (out of character), and then taking two students who made opposite selections of two passages and who *also* produced clearly dissimilar responses to a personality check list, the essential nature of which can be grasped from Table 3.[2]

The speakers then individually tape-recorded the two passages, reading the appropriate in-character passage first. They were allowed to read in whatever way they wished. Tapes suitable to the two conditions of the experiment were prepared by rerecording, in appropriate combination, playbacks of the original recordings.

Following the customary procedure, after listening to each speaker the judges were asked to predict the responses of the speaker to an 18-pair adjective-opposites personality check list. Having done so for each speaker, they were then asked to rate themselves on the same check list. The speakers had already completed the personality check list and their self-ratings served as the subsequent criterion for assessing accuracy. This procedure obviously leaves moot questions of relations between judges' ratings and other criteria for assessing accuracy. It is possible that if judged against a standard personality inventory neither judges' ratings nor self-ratings would appear accurate. However, it is not at all clear, as Allport noted, what constitutes a suitable criterion for a person's true personality. At any rate, for what the judges were asked to do, the authors chose to accept self-ratings as reasonable criteria against which to test the judges' performance. The experimental session lasted about 10 minutes.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF ACCURACY SCORES

Judgment	*M*	*SD*
1st speaker	12.2[a]	2.05
2nd speaker	11.7	1.92
Total (Both speakers)	23.9	3.22

[a] The maximum possible scores for the overall judgment on two speakers and the judgment on one speaker are 36 and 18, respectively.

RESULTS

Since each speaker was judged on 18 items, the maximum accuracy score for individual speakers was 18 and the maximum accuracy score across both speakers was 36 (i.e., these would have been the scores earned by a judge whose predicted ratings were identical with the speakers' self-ratings). On the assumption that judges' predictions should be correct 50% of the time by chance, since the check list required binary choices, *t* tests were performed to determine whether the average observed scores reported in Table 1 differed from chance (i.e.,

---

[2] Dissimilarity was defined as 12 of a possible 18 opposite responses. By using dissimilar (opposite) personalities, it was hoped to exercise some control over social desirability kinds of response sets. In order to be accurate for both speakers, the judge had to perform at least once against the grain of any such predilections.

TABLE 2
ANALYSIS OF VARIANCE OF JUDGES' ACCURACY AND SCALES

Source	*SS*	*df*	*MS*	*F*
Judges (A)	56	77	0.73	1.89*
Scales (B)	138	17	8.12	20.82*
A × B	504	1309	0.39	

* $p < .01$.

from accuracy scores of 9 for single speakers and 18 for both speakers combined). The + − ratios for overall mean accuracy (15.6), for mean accuracy on Speaker A (13.9), and for mean accuracy on Speaker B (12.4) were significant beyond the .001 level, thus providing evidence of accurate judgment.

In the matter of the consistency of judges' accuracy across samples of speech, a product-moment correlation of .40 ($p < .01$) was obtained between judges' accuracy scores on Speaker A and on Speaker B. Thus, judges who were accurate in rating the first speaker tended also to be accurate on the second one, and judges who were inaccurate on the first speaker were inaccurate the second time. While certainly modest, this result does afford comfort for the idea that accuracy of person perception is a general ability, at least within the limits of a given situation.

To explore the effects of open- versus closed-mindedness and of the passages read by the speakers, a 3 × 2 factorial analysis of variance of the randomized-blocks design was performed on the total or combined accuracy scores. It had been assumed, of course, from the Markel-Roblin work that passage variation alone would be without effect. The main effects and interaction of these two factors were not significant.

TABLE 3
ORDER OF DIFFICULTY FOR THE EIGHTEEN SCALES

Scale	% judges who made mistakes
Forceful-Gentle	13.4
Assertive-Reserved	15.4
Realistic-Idealistic	16.0
Bold-Cautious	18.6
Quiet-Talkative	20.5
Youthful-Mature	23.1
Sensual-Intellectual	27.6
Critical-Supporting	28.2
Gregarious-Detached	30.8
Patient-Impatient	32.6
Emotional-Unemotional	34.6
Religious-Nonreligious	38.5
Simple-Complex	41.3
Thick-skinned-Sensitive	43.0
Competitive-Cooperative	46.1
Serious-Humorous	48.6
Orderly-Casual	51.9
Agile-Slow	74.4

Passage variation was again tested, this time in company with role variations (i.e., in character versus out of character), via a split-plot design analysis of variance of the combined accuracy scores. Once again, no significant main effects or interactions were found. Finally, *t* tests of mean accuracy scores for males and females yielded no significant differences.

Turning now to questions of individual differences in accuracy and to differences among the scales in eliciting accurate judgments, the results of a Judges × Scales (rating scales or traits) analysis of variance performed on the combined accuracy scores can be found in Table 2. It will be seen that both main effects were significant at the .01 level. Clearly, then, persons do vary in their abilities to render accurate judgments of others, and, moreover, it is also apparent that some attributes are easier to judge accurately than are others. It is of special interest, perhaps, that judges and attributes do not interact. Judges seem to be good or poor and attributes difficult or easy, but independently of one another.

Table 3 lists the bipolar traits from the personality check list in order of their difficulty. As the table shows, traits such as forceful-gentle and assertive-reserved elicited accurate responses much more frequently than did such traits as serious-humorous or orderly-casual. The fact that accuracy varies with the trait being judged as well as with the judge is consistent with the findings of Bronfenbrenner and his co-workers (1958) and, as Roger Brown has suggested (1965), could underlie the inconsistent findings in the accuracy literature.

Finally, a similarity score between a speaker and a judge was obtained by counting the number of common traits checked by them. These similarity scores then were correlated with judges' accuracy scores for the appropriate speaker. Neither correlation was significantly different from zero ($r = -.12$ for Speaker A and $r = .05$ for Speaker B).

DISCUSSION AND CONCLUSIONS

In summary, then, our findings do provide evidence of accurate judgment of personal attributes from speech. Also, some evidence of consistency of individual performance across samples of speech expressive of quite dissimilar personalities was adduced along with demonstration of significant individual differences in accuracy. Both of these findings are consonant with interpretations of accuracy of judgments in terms of individual abilities (e.g., Bronfenbrenner et al., 1958; Cline & Richards, 1960, 1961). That is to say, in this research the judgments of subjects exhibited variations and consistencies typical of psychological abilities. At the same time, however, the authors' data clearly revealed variations in accuracy as a function of the attribute being judged. As best as can be judged, accuracy from voice cues seems greater for affective-conative attributes than it is for more behavioral-physical ones. One should keep in mind, however, that the dimensions of personality comprised by the authors' check list were relatively simple. It appears that effective descriptions of personality can be accomplished using adjectival formats, but the richness of description possible in this research was limited, and so, therefore, must be generalizations about variations in accuracy across personality dimensions.

Turning to the negative side of the ledger, the authors found accuracy of judgment to be unaffected by whether the judge was open- or closed-minded or by whether the speaker performed in- or out-of-character verbal parts. Nor was any relation between accuracy and interpersonal similarity found.

The most noteworthy findings to emerge from this research are those that support the idea that stable cues to personality are carried by general voice "qualities" *independently of the lexical content* of speech. The fact that listeners could judge personality accurately from speech samples, but that passage content had no independent effect on accuracy suggests either that speech content tends to be no more than redundant with voice quality or simply irrelevant to the judgmental task. Exactly what qualities of voice carry cues to personality and how they operate await specific explication; one may surmise, though, that timbre, inflection, and stress are involved (see Kramer, 1963).

REFERENCES

ALLPORT, G. W. *Pattern and growth in personality.* New York: Holt, Rinehart & Winston, 1961.

BRONFENBRENNER, U., HARDING, A. H., & GALLWEY, M. O. The measurement of skill in social perception. In D. C. McClelland, A. L. Baldwin, U. Bronfenbrenner, & F. L. Strodtbeck (Eds.), *Talent and society.* New York: Van Nostrand, 1958. Pp. 29–111.

BROWN, R. *Social psychology.* New York: Free Press, 1965.

CLINE, V. B., & RICHARDS, J. M. Accuracy of interpersonal perception: A general trait? *Journal of Abnormal and Social Psychology,* 1960, **60**, 1–7.

CLINE, V. B., & RICHARDS, J. M. The generality of accuracy of interpersonal perception. *Journal of Abnormal and Social Psychology,* 1961, **62**, 446–449.

KRAMER, E. Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin,* 1963, **60**, 408–420.

MARKEL, N., & ROBLIN, G. The effect of content and sex-of-judge on judgment of personality from voice. *International Journal of Social Psychiatry,* 1965, **11**, 295–300.

(Received July 14, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 454-457

INCREASE IN PSYCHOLOGICAL ANOMIE AS A RESULT OF RADICAL AND UNDESIRABLE CHANGE EXPECTANCY[1]

GUNARS REIMANIS

Division of Social Sciences, Corning Community College[2]

Support was shown for hypotheses derived from anomie theory that the anomic state of the individual increases in the presence of events leading to expectation of radical and undesirable change in his life, and that it decreases once such events are removed. Significant mean differences in anomie were found when data were collected prior to an order to close a Veterans Administration domiciliary, right after the order, and after the order was rescinded. Ss were domiciliated war veterans, employees at the affected domiciliary, members of a community whose economy largely depended on the domiciliary, and residents of a nonaffected community.

In past research the study of the psychological anomie using the Srole Anomie Scale (Srole, 1956) has been conducted primarily through correlational techniques. The major findings have suggested that (*a*) the roots of anomie may begin early in life through faulty socialization processes (Reimanis, 1965b, 1966); (*b*) anomie varies predictably with a person's value systems as measured by the semantic differential (Reimanis, 1961) and personality factors as measured by the Cattell Sixteen Personality Factor Questionnaire (Reimanis, 1965a); (*c*) certain demographic and psychological variables correlate predictably with anomie (Bell, 1957; Davol & Reimanis, 1959; Reimanis & Davol, 1961); and (*d*) poor life chances may provide an index for predicting anomie (Meier & Bell, 1959).

The present research tested the hypothesis that the occurrence of events giving rise to expectancy of radical undesirable changes in the adult's living pattern leads to an increased individual anomic state. This hypothesis could be derived from several theoretical positions on anomie. It seems, however, that DeGrazia's (1948) theoretical system is most relevant. DeGrazia described anomie as a developmental process which affects the individual whenever radical changes concerning his future take place. During early separation anxiety the individual is seen as being sensitized against any radical change later in life. Any change, such as leaving home for the first time, becoming an adolescent, losing a job, changing residence, involves a period of uncertainty and thus may lead to a temporary state of individual helplessness, normlessness, planlessness, or anomie.

METHOD

The test of the hypothesis was made possible by the announcement of the United States Veterans Administration that a certain Veterans Administration Center would be closed in about 4 months. About 700 domiciliary members with median age of 67 (domiciliary members are war veterans who for varied reasons cannot support themselves in the general community) expected a radical undesirable change, since they would be transferred to other parts of the country which in many cases involved moving away from close friends or relatives in the vicinity. The closest other Veterans Administration domiciliary was at a distance of about 500 miles. An unpublished survey indicated that practically all members preferred to remain at their present domiciliary. About 600 employees at the Veterans Administration Center, which included a hospital, expected a radical change with the possibility of loss of job or change in job at a financial loss and change in residence. In addition, the community with about 6,000 inhabitants at which the Center was located expected a radical change because the closing of the Veterans Administration facility would necessitate a serious economic reorganization. The veterans domiciliary facility had been a major employer for the community for over 80 years, and much of the pension income received by the domiciliary members was being spent in the community.

It was thus predicted that anomie would increase in the domiciliary members, the employees, and the residents of the local community following the Veterans Administration closing announcement.

About 4 months after the closing order, following a series of investigations by congressional subcommittees and other national groups that were set into action primarily by the local leaders' cry for help, the President of the United States announced that the Veterans Administration facility would remain

---

[1] Part of this work was supported by the College Center of the Finger Lakes research grant program. Appreciation is expressed to the Corning Community College students who aided in some phases of data collection, and to the individuals who served as subjects in the study.

[2] Also at Bath Veterans Administration Center, New York.

open for an indefinite period. At this time a second hypothesis was made: namely, anomie will decrease following the rescinding of the closing order.

Subjects

The subjects for the present study were selected from four populations: (*a*) Veterans Administration domiciliary members, (*b*) Veterans Administration employees, (*c*) residents of the affected community, and (*d*) residents of a nearby nonaffected control community with about 20,000 inhabitants and several major industries. An approximation of random-selection procedure of subjects was followed in the study. It was felt that the sampling techniques did not deviate enough from the random technique to preclude the use of normal probability statistics.

For the first sample after the closing order, 130 subjects were selected from the male domiciliary member population on whom anomie data were available from research prior to the closing order. After the rescinding of the closing order 92 of the 130 subjects were still at the domiciliary and they served as subjects in testing the second hypothesis. In the domiciliary sample each subject served as his own control.

The second sample was drawn from the Veterans Administration Center employee population. Anomie data were available on 98 male and 75 female subjects from anonymously conducted research prior to the closing order. After the closing order anonymous questionnaires were sent out to 260 employees. From these, 32 males and 37 females completed and returned the forms. The completion and return rate of the questionnaires decreased from 85% during data collection prior to the closing order to 27% after the closing order. The difference could be understood as a result of the informally observed lower employee morale after the closing order. The employees were not used as subjects after the closing order was rescinded.

The third sample was drawn from the community which received its major economic support from the Veterans Administration Center. After the closing order, data were collected from 99 male and 104 female subjects in a door-to-door canvas through representative streets of the community.[3] After the rescinding of the closing order the same data collection procedure was repeated in the affected community to obtain anomie measures on 44 male and 94 female subjects.

The fourth sample was drawn after the closing order from the nearby control community, economically unaffected by the Veterans Administration installation. Similar door-to-door canvas was carried out in the control community and data were collected from 72 male and 63 female subjects. It is recognized that in the control community private industry was the main economic force and thus may differ from the affected community on other variables not controlled in this study.

TABLE 1

DESCRIPTIVE AND SIGNIFICANCE TEST DATA FOR MEAN DIFFERENCE OF ANOMIE BEFORE, DURING, AND AFTER THREAT OF VETERANS ADMINISTRATION CENTER CLOSING

Population tested	*M* education	*M* anomie	*t* between anomie scores	*N*
Domiciliary members				
Prior to closing order	7.56	8.91		130
After closing order	7.56	9.89	3.26**	130
After rescinding of closing order	7.52	8.74	2.11*	92
VA employees				
Prior to closing order	11.48	7.75		151
After closing order	11.52	9.04	3.23**	69
Control community	12.51	7.36		103
Affected community				
After closing order	12.41	9.31	3.76***	162
After rescinding of closing order	12.68	7.46	4.74***	96

Note.—For domiciliary members *t* values were for matched pairs. Other samples were matched on education before statistical analyses.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Questionnaires

The main part of the questionnaire consisted of five anomie opinion items. The subject expressed agreement or disagreement with these items on a 5-point scale. The response was scored from 0 to 4 with a maximum score of 20. The scale (Reimanis, 1966) was a slightly modified Srole Anomie Scale (Srole, 1956), and was designed to assess essential characteristics of the anomic mode of thought. The questionnaire included an opposite-item anomie scale in all samples but one, as a check for response set. In addition to other less relevant items, questions were included regarding the subject's belief as to whether or not the Veterans Administration facility would be closed and how directly this might affect the respondent.

Previous research (Reimanis & Davol, 1961) has indicated that the anomie scale has construct validity, and at least moderate reliability. Test-retest correlation coefficients for periods of 2 or more years have ranged from .67 to .80. More recently, Struening and Richardson (1965) in a factor analysis of 68 items designed to measure alienation, anomie, and authoritarianism have shown that all five of Srole's anomie items had highest loadings, .54 to .64, on the first factor.

RESULTS AND DISCUSSION

Table 1 summarizes the descriptive and statistical test data comparing the mean anomie scores for the various samples before the closing order, after the closing order, and after the rescinding of the closing order. The findings from the Veterans Administration domiciliary samples, as shown in Table 1, supported both hypotheses. The mean anomie increased significantly following the closing order, and decreased significantly following the rescinding of the closing order. The *t* test computations for the domiciliary members were for matched pairs, since each subject served as his own control.

---

[3] The actual data gathering for this sample was carried out by Raymond Wolfe, State University of New York, College of Geneseo.

With respect to the other samples, subject-by-subject matching was not possible because the data were collected anonymously. The data from these samples showed differences in the distribution of education levels. The means ranged from 11.20 to 12.85. It has been shown in the past that education is an important correlate of anomie (Reimanis & Davol, 1961), and in the present study the correlations between education and anomie ranged between −.22 and −.56. Thus, stratified sampling was used to match the groups by discarding subjects randomly to obtain comparable proportions of cases in each of six education levels for each pair of samples between which *t* values were computed. The means of education, shown in Table 1, are those computed after matching. In testing for mean differences, stratified-sampling variance of mean difference was used for the matched samples (McNemar, 1962, p. 385). Matching was not attempted on age, since chronological age by itself does not seem to relate to anomie (Reimanis & Davol, 1961). In the present samples, excluding the domiciliary member group for which the median age was 67, mean ages ranged between 40.59 and 44.98, and correlations between age and anomie ranged between .00 and .14 (*ns*).

From the *t* value in Table 1 it can be seen that the mean anomie increased significantly for the employee group following the closing order, and that the affected community had a significantly higher mean anomie after the closing order than after the rescinding of the closing order. In the same way the affected community had a significantly higher anomie after the closing order as compared to the control community. These results again supported the predictions that anomie would increase in presence of radical change expectation and decrease with the removal of events giving rise to such expectation.

No significant trends or differences were found between anomie and the indexes measuring the degree to which the subjects felt that they would be affected by the closing of the Veterans Administration facility or their belief that the Veterans Administration installation would or would not close. The lack of relationship could be due to the fact that the data showed little variability on the indexes. Practically every subject in the community indicated that he would be affected somehow by the Veterans Administration closing. If the subject was not directly affected, he usually was involved through his friends and relatives. In the few cases where a subject had no close friends or relatives working at the Veterans Administration Center, his job could be affected through his place of business, or through the fact that real estate values had already gone through a very steep decline.

No consistent differences in anomie were apparent between males and females. The largest difference which was present and significant ($p < .05$) was between the unmatched male and female employees prior to the closing order. The means were 8.63 for males and 7.17 for females. However, their mean education levels were also significantly different (10.24 and 12.45 for the males and females, respectively). The significant mean difference in anomie disappeared when education was controlled.

Concluding, it may be stated that the first hypothesis was supported in this quasiexperimental, realistic research setting, by showing that events giving rise to expectations of radical undesirable change were accompanied by an increase in the anomic state in the individual. It was also shown that such an increase in anomie could be of a temporary nature which decreased with the removal of events giving rise to such expectation. This second finding in particular supported DeGrazia's (1948) view that the anomic state in the individual increases with conditions that upset his usual living pattern and give rise to uncertainty and planlessness, and returns to a previous base line once the uncertainty is removed. Finally, the study has shown that the anomie scale is sensitive enough to measure changes in the individual that can be predicted on basis of anomie theory. Thus, the present study has also shown support for construct validity of the anomie scale.

REFERENCES

BELL, W. Anomie, social isolation and the class structure. *Sociometry*, 1957, 22, 105–116.

DAVOL, S. H., & REIMANIS, G. The role of anomie as a psychological concept. *Journal of Individual Psychology*, 1959, **15**, 215–225.

DEGRAZIA, S. *The political community*. Chicago: University of Chicago Press, 1948.

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1962.

MEIER, D. L., & BELL, W. Anomia and differential access to the achievement of life goals. *American Sociological Review*, 1959, 24, 189–202.

REIMANIS, G. Relationship between anomie and semantic differential ratings of various culturally well-defined concepts. *Newsletter for Research in Psychology*, 1961, 3, 76–82.

REIMANIS, G. Cattell 16 P. F. study in a VA domiciliary. *Newsletter for Research in Psychology*, 1965, 7, 39–42. (a)

REIMANIS, G. Relationship of childhood experience memories to anomie later in life. *Journal of Genetic Psychology*, 1965, **106**, 245–252. (b)

REIMANIS, G. Childhood experience memories and anomie in adults and college students. *Journal of Individual Psychology*, 1966, **22**, 56–64.

REIMANIS, G., & DAVOL, S. H. Correlates and prediction of anomie in a VA domiciliary. *Journal of Social Psychology*, 1961, **55**, 237–244.

SROLE, L. Social integration and certain corollaries: An exploratory study. *American Sociological Review*, 1956, 21, 709–716.

STRUENING, E. L., & RICHARDSON, A. H. A factor analytic exploration of the alienation, anomia and authoritarianism domain. *American Sociological Review*, 1965, **30**, 768–776.

(Received July 18, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 458-463

EMOTIONALITY AND INFORMATION CONCERNING INTERNAL REACTIONS[1]

STUART VALINS

University of North Carolina

It was hypothesized that emotional and unemotional individuals differ with respect to the utilization of internal sensations as cues. In comparison to unemotional Ss, emotional ones were expected to make more use of information concerning their internal reactions. Ss who were psychometrically classified as emotional or unemotional were shown 10 slides of seminude females while hearing a tape recording of sounds that were allegedly their heart beats. To 5 of the slides they heard a marked change in their "heart rates"; to the other 5 they heard no change. It was found that relatively emotional subjects labeled nudes as attractive or unattractive depending upon whether they thought their hearts had reacted. This effect was significantly less marked for unemotional subjects. These individual differences were still apparent 2 mo. later when subjects were allowed to choose photographs of the nudes as rewards.

The present experiment is concerned with individual differences in the labeling of internal sensations. How might we characterize an individual who ignores his internal reactions or who does not utilize them as cues? Considering those experiments in which the labeling process affected emotion-like behavior, we would expect that someone who ignores his internal sensations would not mimic the behavior of an emotional stooge (Schachter & Singer, 1962), would not consider electric shocks to be very painful (Nisbett & Schachter, 1966), and would not like nude females to which he heard his "heart rate" change (Valins, 1966). There would be little reason for this individual to adopt the emotional labels that are available since he would feel little pressure to evaluate his internal sensations. This individual should be relatively unemotional. The opposite would be expected from those individuals who do view their internal sensations as an important source of information and who do utilize these sensations as cues. Their reactions in response to emotional situations should arouse them to search for appropriate explanations or labels. These individuals should be relatively emotional. Such speculations lead us to suspect that individual differences in chronic emotionality are related to differences in the utilization of internal cues. More emotional individuals may feel more pressure to label their internal sensations than do less emotional ones. It is important to note that we are considering a variable that is cognitive and one that can and may be independent of actual internal reactions. Our concern, at the moment, is not with individual differences in physiological reactions but with individual differences in the cognitive effects of these reactions.

The hypothesis tested in this experiment was that individuals who are psychometrically classified as relatively unemotional will make less use of information concerning their internal reactions when evaluating emotional stimuli than will individuals who are classified as emotional. Subjects were shown slides of seminude females while hearing bogus heart sounds. Half of each emotionality group heard their "heart rates" increase to some slides and not change to others. The remaining subjects heard the bogus heart rate decrease to some slides and not change to others. It has been shown that subjects use this information when evaluating the nudes in that they like best those nudes to which they hear a change in the bogus heart rate (Valins, 1966). However, if unemotional types ignore their internal sensations or do not utilize them as cues, they should not be influenced by the feedback when evaluating the nudes. In comparison to the more emotional subjects, the less emotional ones should consider the nudes accompanied by a change in the bogus heart rate as similar in attractiveness to those nudes not so accompanied.

Inclusion of both feedback manipulations in the design of this experiment constituted a replication of previous research. Although it is intuitively reasonable that subjects would interpret an increase in heart rate to a slide a of seminude female as indicating attraction, it is not so obvious that they would similarly interpret a decrease in heart rate. Replication of this latter effect is thus desirable.

---

[1] This research was facilitated by a faculty research grant from the University of North Carolina School of Medicine, and by funds provided by the Foundations' Fund for Research in Psychiatry. Appreciation is extended to Neil Carson and Douglas Gouge for their able assistance during the initial phase of the experiment.

METHOD

Subject Selection

Emotionality was defined with respect to two paper-and-pencil questionnaires. The prime selection device, Lykken's questionnaire (Lykken, 1957), is in a forced-choice format and consists of 33 items that describe two rather unpleasant activities. One alternative is unpleasant because it is distasteful or tedious, while the other is embarrassing or anxiety arousing. The latter alternatives will presumably be avoided or not preferred by those individuals who are most subject to the emotions of anxiety or fear. Subjects who have been defined as emotional in this manner, in comparison to those defined as unemotional, have been found to be: (*a*) more emotional by clinicians (Lykken, 1957), (*b*) good learners on a shock-avoidance task in which performance is facilitated by anxiety or fear (Schachter & Latané, 1964), (*c*) more upset at the scene of an air disaster (Latané & Wheeler, 1966), and (*d*) more nervous as indicated by their scores on a questionnaire to be discussed below.

Lykken's questionnaire was administered to approximately 1800 male undergraduates at freshman orientation meetings. Those undergraduates who chose 8 or less of the frightening alternatives (8%) and those who chose 20 or more (9%) were selected as potential subjects. The scores of these subjects on a second questionnaire were used to further select the experimental population. This questionnaire (Valins, 1963) consisted of 15 items, each depicting a situation previously judged to arouse some degree of nervousness (e.g., going to the hospital for an operation). The subject circled 1 of 5 points next to each item to indicate how nervous he would be in that situation, and those points circled were summed to arrive at a nervousness score. Of those subjects classified as unemotional using Lykken's questionnaire (score of 20 or more), 66% had nervousness scores that were below the median score (29.5) of all potential subjects, and of those subjects classified as emotional, 69% had nervousness scores that were above the median ($\chi^2 > 24$, $df = 1$, $p < .001$).[2] Subjects were recruited for the present experiment from those who scored 20 or more on Lykken's questionnaire and who also scored 29 or less on the nervousness questionnaire (low emotional subjects—LE) and subjects scoring respectively 8 or less and also more than 29 on these questionnaires (high emotional subjects—HE). As an additional check on whether subjects so selected differed on an anxiety or emotionality dimension, their scores on the *A* or anxiety factor of the MMPI were compared. Emotional subjects who participated in the present experiment scored higher on the anxiety factor than did unemotional subjects (18.23 versus 9.60, $t = 4.76$ $p < .001$).

Manipulation of Information Concerning Internal Reactions

The procedure that was employed has been presented in detail elsewhere (Valins, 1966). Subjects were told that the experiment was concerned with heart-rate reactions to sexually oriented stimuli. These reactions were to be measured in response to 10 slides of seminude females. Due to an alleged inadequacy of instrumentation, the subjects were to actually hear their hearts beating but were told to ignore it. Heart rate was presumably being measured using a microphone that was taped to the subject's chest, and an amplifier and signal tracer that reproduced the heart sounds. The subjects thought that these "heart sounds" were being recorded on a tape recorder. The sounds, of course, were actually prerecorded and were played back through the speaker of the signal tracer.

Ten color slides, which were made from photographs published by *Playboy* magazine, were each projected for 15 seconds at 1-minute intervals through a one-way mirror. The slide presentation was coordinated with the tape recording and began with the third minute. Using this procedure, the subjects heard a marked change in the bogus heart rate to five of the slides but not to the other five. After the tenth slide, the subjects heard their "hearts" beating normally for three minutes. The slides were then repeated in the same manner with subjects hearing the identical reactions. To control for variations in the attractiveness of the nudes, the slide order was systematically rotated within conditions. Twenty male subjects were assigned to each of the following conditions: high emotional–heart rate increase, high emotional–heart rate decrease, low emotional–heart rate increase, and low emotional–heart rate decrease.

Attractiveness Measures

To assess the degree to which the feedback was utilized when evaluating the nudes, several measures of the attractiveness of each nude were obtained from all subjects.

Slide ratings. On the pretext that the experimenter wished to reduce the number of slides to the most attractive ones, the slides were shown a third time (with no feedback), and the subjects rated the attractiveness of each girl, using a 100-point scale.

Photograph choices. As an additional reward for participation (the subjects were paid $3.00 for this experimental session and another to be discussed below) the subject was told he could take 5 of the 10 photographs from which the slides had been made. Copies of his choices, presumably donated by the publisher, would be given to him with his pay check.

---

[2] Reported *p* values are based on two-tailed tests.

Ratings of photograph details. These ratings were not obtained in the original experiment. They were included in this one to determine whether the evaluation of specific features of the nudes would be affected by the feedback. After the subject chose the five photographs that he wanted, he was told that one of the experimenters was also interested in "what about these photographs is good or bad." The subjects were asked to rate all 10 photographs on the basis of how "nice" the breasts were, followed by ratings of the face, hair, and general pose.

Delayed photograph choices. About 6 weeks after participating in the first experimental session, the subjects received a letter requesting them to make an appointment for a second session which they had previously agreed to. They were seen during this session by another experimenter who, after introducing himself, first gave the subject a pamphlet to read and a questionnaire to complete. This took about 15 minutes and, since it was irrelevant to the purpose of the present experiment, will not be discussed. After the subject completed the questionnaire, the experimenter said that he would pay him before the second part of the experiment began. The subject was given his pay check and also told that "rather than looking over my records to see which photographs you wanted, take the five that you want and I'll go get some copies." The subject was given 10 photographs, and the experimenter hurried out of the room, ostensibly to get copies. Upon returning, he explained that the experiment was over and questioned the subject regarding his reactions.

RESULTS

Postexperimental questioning revealed that all but two subjects accepted the bogus heart beats as veridical. These two subjects, one of which was an LE and one an HE, were in the heart-rate increase condition. To retain an equal number of subjects in each condition, the data of these two were included in all analyses. This is a conservative procedure, since on all measures the data indicated that the LE subject had been more affected by the feedback than the HE subject. Exclusion of these data would have operated to strengthen the major results and conclusions.

Slide Ratings

In Table 1 are presented the mean attractiveness ratings, for each experimental condition, of the slides to which the subjects heard a change in the bogus heart rate (reinforced) and the slides to which they heard no change (nonreinforced). It will be noted first that the reinforced nudes were considered more attractive than the nonreinforced ones ($F = 72.15, p < .01$), and that this effect was somewhat stronger with the increase feedback than with the decrease feedback (F for Reinforcement × Feedback interaction $= 6.12$, $p < .05$).[3] However, the decrease feedback did have a significant effect. Seventy-two percent of these subjects rated the reinforced nudes as more attractive than the nonreinforced ones ($z = 2.69$, $p < .01$). These data replicate the results of previous research. Information concerning internal reactions does affect the evaluation of emotional stimuli.

With regard to our primary interest, it was predicted that relative to more emotional subjects, less emotional ones would not utilize the bogus feedback when evaluating the nudes. These latter subjects should consider the two types of photographs as similar in attractiveness. As can be seen in Table 1, the bogus feedback did not equally affect the evaluations of the LE and HE subjects. The difference in the attractiveness ratings of the reinforced and nonreinforced nudes is greater for the HE subjects than for the LE subjects (F for Reinforcement × Emotionality interaction $= 7.61$, $p < .01$). The more emotional subjects utilized the information concerning their internal reactions to a greater extent than did the less emotional subjects.[4]

Photograph Choices

Toward the end of the first experimental session the subjects were told that they could have five of the photographs from which the slides had been made. Analysis of the photographs that were chosen provides another indication of the differential effects of the bogus feedback. The mean number of reinforced and nonreinforced photographs that were chosen is presented in Table 1. In accord with the analysis

---

[3] All F values result from $2 \times 2 \times 2$ (Emotionality × Feedback × Reinforcement) analyses of variance, with $df = 1/76$.

[4] The F values for the main effect of emotionality and for the interaction of Emotionality × Feedback × Reinforcement did not approach significance. However, the interaction of Emotionality × Feedback was significant ($F = 4.20$, $p < .05$). This interaction indicates that with the increase feedback the more emotional subjects tended to give higher attractiveness ratings than the less emotional subjects, and that with the decrease feedback they gave lower ratings. This significant interaction and the nonsignificant main effect of emotionality indicate that the HE and LE subjects did not differ with respect to the magnitude of their attractiveness ratings.

TABLE 1

MEAN SLIDE ATTRACTIVENESS RATINGS AND PHOTOGRAPH CHOICES

Slides	Conditions			
	HE		LE	
	Increase	Decrease	Increase	Decrease
Reinforced	79.99 (3.70)	66.50 (3.25)	69.77 (3.20)	66.39 (2.75)
Nonreinforced	56.65 (1.30)	51.78 (1.75)	56.01 (1.80)	60.75 (2.25)
Difference	+23.34 (+2.40)	+14.72 (+1.50)	+13.76 (+1.40)	+5.64 (+.50)

Note.—Numbers in parentheses are the mean number of photographs chosen as rewards.

of slide ratings, more of these choices were of reinforced photographs ($F = 49.47, p < .01$), and a stronger effect was obtained with the increase feedback than with the decrease feedback ($F = 4.76, p < .05$). Again, however, the decrease feedback did have a significant effect. Seventy percent of these subjects chose more reinforced photographs than nonreinforced ones ($z = 2.37$, $p < .02$).

It is also apparent that the choices of the more emotional subjects were more effected by the feedback than were the choices of the less emotional ones. In comparison with the less emotional subjects, the more emotional subjects chose more reinforced photographs than nonreinforced ones ($F = 5.94$, $p < .05$).[5] This effect has also withstood a delay of approximately 67 days. During the second experimental session the subjects were again told that they could have five photographs as a reward for participating in the experiment. It should be emphasized that although no attempt was made to dissociate the two experimental sessions, the subjects made these delayed choices on the basis of the photographs that they liked best at the moment. Although many of them were aware of the consistency of their choices, and some suspected that the experimenter would compare the immediate and delayed ones, virtually all of the subjects insisted that they liked the five photographs that they had chosen more than the five they had not. Of the 33 HE subjects who returned for the second session, 28 chose more reinforced photographs than nonreinforced ones; of the 30 LE subjects who returned, 18 chose more reinforced photographs. This delayed effect ($\chi^2 = 3.74$, $df = 1$, $p < .06$) is probably a conservative estimate, however, because those HE subjects who did not return previously chose more reinforced photographs than the HE subjects who did return, and those LE subjects who did not return previously chose less reinforced photographs than the LE subjects who did return. That internal information is of greater importance to more emotional individuals than to less emotional ones is strongly supported by these delayed data.

Ratings of Photograph Details

It is evident from the data presented so far that subjects like best those nudes to which they think their hearts have reacted. It is also of interest, however, to determine whether this effect is manifested in the subjects' attitudes toward specific features of the nudes. If similar effects are noted, they would suggest that the subjects distort various features of the nudes to explain their presumed heart-rate reactions, and that these distortions may serve to justify considering a nude as attractive or unattractive. At the end of the first experimental session, the subjects rated each photograph (using a 100-point scale) on the basis of "how nice are the girl's breasts, face, hair, and general pose." The mean differences between these ratings of the reinforced and nonreinforced nudes are presented in Table 2. These data are consistent with the slide ratings and photograph choices. The breasts, faces, and poses of the reinforced nudes were rated as nicer than those of the nonreinforced ones (all F values > 12.46, $p < .01$), and a similar trend is apparent for the ratings of the nudes' hair ($F = 3.60$, $p < .10$). These effects were also somewhat greater with the increase feedback than with the decrease feedback (F values for faces and poses > 4.31, $p < .05$).

These ratings are not only dependent upon the bogus feedback but are also dependent upon whether a subject is more emotional or less emotional. The difference between the ratings of the reinforced and nonreinforced nudes is greater for the more emotional subjects than for the

---

[5] None of the remaining F values for this analysis were significant.

TABLE 2

PHOTOGRAPH DETAILS: MEAN RATINGS OF REINFORCED PHOTOGRAPHS MINUS MEAN RATINGS OF NONREINFORCED PHOTOGRAPHS

Details	Conditions			
	HE		LE	
	Increase	Decrease	Increase	Decrease
Breasts	+8.21	+5.92	+ .85	+1.04
Faces	+7.69	+3.08	+2.41	− .46
Hair	+4.61	+4.52	+2.90	− .30
Poses	+15.47	+7.76	+10.60	+ .53

less emotional ones for the ratings of breasts ($F = 11.30$, $p < .01$), faces ($F = 5.99$, $p < .05$), and poses ($F = 3.30$, $p < .10$). Once again, the data indicate that the more emotional subjects have been more affected by the bogus feedback than have the less emotional subjects.[6]

Control Comparisons

Two additional groups of subjects were used to determine whether the tape-recorded sounds alone, apart from their meaning, had a differential effect on the emotional and unemotional subjects. These subjects heard the heart-rate increase tape recording and, except for thinking that the recording was of "extraneous sounds,"[7] were exposed to the same first-session procedures as the experimental subjects. No significant differences were found on any measures between control subjects classified as emotional and those classified as unemotional ($N = 9$ for each group). In fact, for five of the six comparisons, the differences between the control groups were in a direction opposite to the differences between the experimental groups. Thus, the latter differences are due to the sounds functioning as information about internal reactions.

DISCUSSION

The results of this experiment are in agreement with theoretical expectations. Individuals do utilize information concerning their internal reactions, and the degree to which this is done can be predicted by knowing whether an individual is more emotional or less emotional. These individual differences highlight the importance of cognitive processes in mediating emotional response. Although physiological reactions may indeed be important determinants of emotion, perhaps more important is whether an individual feels it necessary to have a label or explanation for his internal state. It is conceivable, for example, that an individual who is characterized by marked physiological reactivity could be relatively unemotional, whereas someone who is less reactive could be relatively emotional. The former might ignore his internal reactions and the latter might experience much pressure to label and evaluate them.

Just such an explanation has been advanced to account for the emotional flatness of the criminal sociopath. Contrary to predictions derived from a cognitive and physiological theory of emotion, Schachter and Latané (1964) have found that the chronically unemotional criminal shows indications of sympathetic hyperreactivity. Thus, the absence of internal cues could not account for his lack of emotion. However, the absence of pressures to evaluate these cues could explain the data. Regardless of how or why they react internally, sociopaths may be unemotional *because* they ignore or do not utilize as cues whatever internal sensations they do experience.

This explanation of the behavior of the unemotional criminal is supported by these results if one can consider the subject-selection criteria as comparable to those of Schachter and Latané. In the present experiment, emotionality was primarily defined by scores on Lykken's questionnaire. As reviewed above, Lykken found that these scores were related to clinical nominations of criminals as emotional or unemotional, and the questionnaire was one of several criteria employed by Schachter and Latané to select their sociopaths. Experiments that have used Lykken's questionnaire with undergraduate populations have also found similar physiological differences between emotional and unemotional individuals (Valins, 1963, 1967). Furthermore, there is collateral evidence indicating that our unemotional subjects were sociopathic-like. One of the characteristics of the sociopath is his lack of responsibility. He does not seem to be concerned with the consequences of his actions. Such irresponsibility was evident in the appointment-keeping behavior of the LE subjects in the present experiment. Although no records were kept of this behavior for the first experimental session, they were for the second session. Of the

---

[6] Of the remaining F values for these four analyses, the only ones that approached significance were the four interactions of Emotionality × Feedback (all F values > 3.17, $p < .10$). These parallel the similar interaction manifested with the slide ratings.

[7] See Valins (1966) for a more complete description of these control conditions.

35 LE subjects who called and arranged for a second session appointment, 13 of them (37%) failed to show up; of the 33 HE subjects who called, only one (3%) did not keep his appointment.[8] This greater irresponsibility on the part of the unemotional subjects ($\chi^2 = 10.10$, $df = 1$, $p < .01$) further indicates that the subjects were selected on a dimension similar to that of Schachter and Latané. It would thus appear reasonable to conclude that certain kinds of unemotional individuals are sympathetically hyperreactive, but also ignore and do not label their reactions. Why they are hyperreactive and why they ignore their reactions are problems for further investigation.

---

[8] Eight LE subjects and the one HE subject returned after being scheduled for another appointment.

REFERENCES

LATANÉ, B., & WHEELER, L. Emotionality and reactions to disaster. *Journal of Experimental Social Psychology,* 1966, 1, 95–102.

LYKKEN, D. T. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology,* 1957, 55, 6–10.

NISBETT, R., & SCHACHTER, S. Cognitive manipulation of pain. *Journal of Experimental Social Psychology,* 1966, 2, 227–236.

SCHACHTER, S., & LATANÉ, B. Crime, cognition, and the autonomic nervous system. In D. Levine (Ed.), *Nebraska Symposium on Motivation: 1964.* Lincoln: University of Nebraska Press, 1964.

SCHACHTER, S., & SINGER, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, 69, 379–399.

VALINS, S. Psychopathy and physiological reactivity under stress. Unpublished master's thesis, Columbia University, 1963.

VALINS, S. Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology,* 1966, 4, 400–408.

VALINS, S. Emotionality and autonomic reactivity. *Journal of Experimental Research in Personality,* 1967, 2, 41–48.

(Received July 18, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 464–466

RECOGNITION OF ONE'S OWN VOICE[1]

CLYDE ROUSEY AND PHILIP S. HOLZMAN

The Menninger Foundation

In a study of the effects on voice recognition of (a) length of utterance, (b) delay in playback, and (c) familiarity with voice, only familiarity appeared to be a significant influence on correctness of recognition. Within the limits of 5 sec. of utterance and 3 mo. delay between recording and playback, about 38% of Ss with only moderate experience in hearing their own voices and with no professional interest in their own voice correctly recognized their own voices. When the factor of interest and familiarity are added, correct recognition can be significantly increased.

Previous studies have demonstrated that when a person listens to a recording of his own voice he experiences a disruptive impact which is discernible not only in verbal responses but also in psychophysiological reactions (Holzman & Rousey, 1966; Holzman, Rousey, & Snyder, 1966). In one of those studies (Holzman et al., 1966), psychophysiological responses were monitored while the subject listened to 20 voices saying the same phrase, only 1 of which was the subject's own voice. Own voice produced significantly higher psychophysiological activation than did other voices, a trend that was maintained whether or not subjects recognized their own voices. A forced-choice recognition test showed that only 15 of the 39 subjects (38%) recognized their own voices. In an earlier study, Wolff (1943) reported that only 10.5% of his subjects correctly identified their own voices on immediate playback, but none correctly identified his own voice after an interval of a few weeks. The high rate of nonrecognition in their and Wolff's studies intrigued the authors and they wanted to determine some of the main factors influencing own-voice recognition. The present experiment studies the effect of three variables on recognition of one's voice. These variables were selected on a priori grounds as accounting for most of the difficulty in recognizing own voice. The variables are (*a*) the length of the utterance, (*b*) the time interval between recording and listening, and (*c*) familiarity with one's recorded voice.

EXPERIMENT I

Method

Subjects. Two hundred ninety-six subjects were tested. They were recruited from several civic clubs, church auxiliaries, and national sorority alumnae groups. All subjects were middle-class midwesterners with approximately the same regional accent.

Procedure. The task required subjects to tape-record a short utterance. Twenty subjects were recorded on a single tape. Then subjects listened to the tape with the 20 voices all saying the same phrase, and they were instructed to mark in a booklet whether each voice was his own or not.

The authors used two lengths of utterance: 1 second, and 5 seconds.[2] There were three time intervals between recording and playback: 30 minutes, 1 week, and 3 months. Subjects were haphazardly assigned to one of the six conditions for the voice-recognition test. A portable Uher Model 4000 S tape recorder was used. Segments of tape, the lengths of which corresponded to the lengths of the 1-second or 5-second utterances with 5 seconds of silence between the segments, were previously numbered. After the subjects had recorded their voices the tapes were cut, rearranged, and respliced in a prearranged random order. This was done to prevent the subjects from responding to the order in which they had recorded their voices, for some may have remembered that they recorded before or after a particular subject. The instructions were:

> We want to see if you can recognize your own voice. We are going to play for you a series of 20 voices. Your own voice may appear in this series once, many times, or not at all. After each exposure, mark in your booklet on the appropriate page whether you believe that the voice you hear is your own. If you think it is, check the box adjacent to "Yes"; if you think it is not, check

---

[1] This investigation was supported by Grant MH 07962 from the National Institute of Mental Health.

The authors acknowledge the help of a number of people. Winifred Siegel and Linda Morton assisted the authors in testing the subjects. Lolafaye Coyne supervised the statistical treatment of the data. Paul Winders, chief engineer of WIBW-TV, Topeka, Kansas, made valuable technical suggestions for recording and editing the magnetic tapes. Hazel Bruce bore all of the chores of preparing the manuscript.

[2] The 1-second sample was, "a long flowing beard." The 5-second sample was, "He dresses himself in an ancient black frock coat usually minus several buttons." These excerpts were taken from a passage entitled, "My Grandfather (Van Riper, 1954)."

TABLE 1

DISTRIBUTION OF SUBJECTS WHO DID AND DID NOT RECOGNIZE THEIR OWN RECORDINGS UNDER VARYING LENGTHS OF PASSAGE AND TIME DELAYS BETWEEN RECORDING AND LISTENING

Period of playback delay	Length of utterance							
	1 Sec.				5 Sec.			
	Recognize		Nonrecognize		Recognize		Nonrecognize	
	N	%	*N*	%	*N*	%	*N*	%
Immediate playback	26	46	31	54	20	34	39	66
1 wk. delay in playback	20	36	35	64	14	39	22	61
3 mo. delay in playback	17	33	35	67	15	41	22	59
Total	63	38	101	62	49	37	83	63

"No." Now here is a practice trial. [Practice stimulus given.]

Remember, your own voice may appear anywhere in the series; it may appear once, several times, or not at all. Please don't talk while the experiment is going on. It is very important that silence be maintained because any comment or exclamation may give hints to the others. So please don't laugh or whisper or groan or so on during the experiment.

Since people should not know anything about these studies before they serve as subjects, it is extremely important that you do not discuss what you do here with anyone at any time. We will, of course, tell you about the experiment in great detail after we have finished gathering all of the data.

Now we are ready to start. Are there any questions?

The authors also asked half of the subjects to indicate the frequency with which they heard their own recorded voices. The choices were (*a*) at least once a week, (*b*) at least once a month to less than once a year, and (*c*) never.

The number of subjects in each of the conditions are shown in Table 1.

TABLE 2

DISTRIBUTION OF CORRECT AND INCORRECT RECOGNITION OF OWN VOICE BY 169 SUBJECTS WITH DIFFERENT DEGREES OF EXPERIENCE IN HEARING THEIR OWN VOICES

Degree of experience	Voice recognized		Voice not recognized	
	N	%	*N*	%
More than once a wk.	5	83	1	17
Once a mo. to once a yr.	46	34	88	66
Never	8	28	21	72

RESULTS

Only results pertaining to recognition of voice are reported here. Table 1 shows the numbers and percentages of subjects who did and did not recognize their own voices under each of the six conditions. A multidimensional chi-square analysis (Sutcliffe, 1957) reveals no significant differences between any of the conditions in voice recognition. Thus, the number of people who did not recognize their own voices is rather constant whether subjects heard a 1- or 5-second sample of their own voices and whether the playback was within 30 minutes, 1 week, or 1 month after recording. When one combines the results of all six conditions, 38% of the subjects correctly recognized their own voices. This statistic is identical to the 38% figure found in the authors' previous experiment (Holzman et al., 1966).

Table 2 shows the effects of experience with listening to one's voice on correct recognition. A multidimensional chi-square analysis shows a significant interaction between frequency of listening to own voice and correct recognition ($p < .05$). This finding led to Experiment II, a more systematic exploration of the effect of familiarity with one's own recorded voice on recognition.

EXPERIMENT II

To determine the influence of frequency of listening to and familiarity with one's own voice on voice recognition, the authors tested two additional groups who presumably would be more familiar with their own voices than the subjects in Experiment I. One group consisted of 28 professional radio announcers who hear their own voices several times each day. The second group consisted of 20 third-year ministerial students at a Methodist seminary. Their seminar curriculum in homiletics requires close attention to their speaking voices, although the frequency with which they listen to their recorded voices is not significantly greater than that of the subjects in Experiment I. The procedure for this experiment was identical to that in Experiment I except that the authors used for the radio announcers a 1-second utterance with a 1-week delay in the playback, and for the ministerial students, a 1-second utterance with a 30-minute delay in playback.

RESULTS

The recognition results appear in Table 3.

TABLE 3

FREQUENCY OF RECOGNITION AND NONRECOGNITION OF OWN VOICE BY RADIO ANNOUNCERS AND MINISTERIAL STUDENTS

Group	Voice recognized		Voice not recognized	
	N	%	*N*	%
Radio announcers	28	100	0	0
Ministerial students	13	65	7	35

Clearly the rate of recognition is higher in the radio announcers and ministerial students than in those subjects used in Experiment I. All of the radio announcers correctly recognized their own voices, and 65% of the ministerial students correctly identified their own voices.

REFERENCES

HOLZMAN, P. S., & ROUSEY, C. The voice as a percept. *Journal of Personality and Social Psychology,* 1966, 4, 79–86.

HOLZMAN, P. S., ROUSEY, C., & SNYDER, C. On listening to one's own voice: Effects on psychophysiological responses and free associations. *Journal of Personality and Social Psychology,* 1966, 4, 432–441.

SUTCLIFF, J. P. A general method of analysis of frequency of data for multiple classification designs. *Psychological Bulletin,* 1957, 54, 134–137.

VAN RIPER, C. *Speech correction: Principles and methods.* (3rd ed.) Englewood Cliffs, N. J: Prentice-Hall, 1954. Pp. 178–179.

WOLFF, W. *The expression of personality.* New York: Harper, 1943.

(Received July 21, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 467-471

WILL AN OBSERVER ADVISE HIGHER RISK TAKING AFTER HEARING A DISCUSSION OF THE DECISION PROBLEM? [1]

HELMUT LAMM
Wirtschaftshochschule Mannheim

If debate over a decision results in higher risk taking on the part of group members, will physically separate observers be affected in the same way? In this study a series of hypothetical decision situations—the Choice Dilemmas task devised by Wallach and Kogan—was used to measure S's risk-taking propensity. The group members as well as the noninteracting "viewers" (placed behind a one-way mirror) and "listeners" (in a separate room with a loudspeaker) were found to shift toward higher risk. The evidence supports the theory of information exchange as the explanation of the risky shift rather than the theory of responsibility diffusion.

It has been shown in a number of studies that groups tend to take higher risks than individuals. Risk-taking level has in most cases been measured through a Choice Dilemmas instrument (Wallach & Kogan, 1959) containing brief descriptions of 12 hypothetical decision problems. The subject is asked to indicate the minimum odds of success he would demand before recommending the choice of the more attractive but riskier alternative among two possible courses of action.

Brown (1965) and Kogan and Wallach (1967) have reviewed the relevant research and advanced differing explanatory approaches. According to the latter authors, the source of the risky shift lies in a process of responsibility diffusion. This concept was first proposed by Wallach, Kogan, and Bem (1962), when they reported a confirmation of the original finding by Stoner (1961) of the group risky shift. In a more recent study, Wallach and Kogan (1965) found no risky shift when subjects were asked to disclose their positions (Choice Dilemmas responses), and reach consensus through repeated rounds of public balloting without discussion. Shifts occurred in the discussion condition even when no group decision (consensus) was required. This finding necessitated a broader conceptualization of responsibility diffusion. Whereas first (Wallach et al., 1962) the idea was that the subject felt his responsibility reduced because the decision was no more his alone, it is now the affective bonds engendered through the give and take of verbal interaction that "enables him to feel less to blame when he entertains the failure of a risky decision [Wallach & Kogan, 1965, p. 18]."

The notion of responsibility diffusion is absent in the value theory proposed by Brown (1965) to explain the risky-shift phenomenon. Drawing on unpublished data from Hinds (1962)—that individuals tend to see themselves as high risk takers relative to unspecified others—and Nordhøy (1962)—that statements favoring risk outnumber those favoring caution in the group discussion—Brown argues that the exchange of information in the group leads to the realization of risk as a value, thus causing a shift toward higher risk taking. Recent research by Rabow, Fowler, Bradford, Hofeller, and Shibuya (1966), and Teger and Pruitt (1966) support Brown's theory. Their findings cannot be explained by the diffusion of responsibility concept. Teger and Pruitt found that repeated information exchange among subjects regarding their Choice Dilemmas responses—with discussion not allowed and consensus not required—resulted in a risky shift.

The present study was designed as another test of the two explanatory approaches. Will a nonparticipating and isolated witness remain unaffected by the discussion or will he shift toward higher risk along with the group? A subject in such a condition is not taking part in any group decision nor in the verbal give and take which Wallach and Kogan (1965) concluded to be necessary and sufficient for the risky shift to occur. Therefore, if responsibility diffusion is the primary source of the risky shift, the noninteracting observer should not exhibit the effect. The central concept of the Kogan-Wallach theory would have to be further broadened to accommodate a contrary finding. Brown's theory, on

---

[1] Financial support for this study came from the Office of Naval Research, Contract No. 1858-(12), Harold M. Schroder, prinicipal investigator. The author thanks H. M. Schroder, J. L. Kennedy, N. Kogan, and B. L. Koslin for their generous advice and helpful criticisms.

This article is based on part of the author's doctoral dissertation at Princeton University.

the other hand, predicts a risky shift for a nonparticipating witness just as well as for the group members, since the information exchange—disclosure of individual positions and risk-relevant argumentation—is operative in either condition.

METHOD

Subjects

The 78 subjects, paid volunteers, were male students at Princeton University. Base-line scores on the Choice Dilemmas instrument were obtained through a testing session a few weeks prior to the experiment. In assigning subjects to the three experimental conditions (groups, viewers, and listeners), care was taken that they did not differ with respect to these preexperimental scores.

In order to be selected for the experiment, subjects had to be extreme (low or high) in conceptual complexity as indicated by their scores on the Paragraph Completion Inventory (Schroder, Driver, & Streufert, 1967). The 16 experimental groups were homogeneously composed so that the discussions of the simple and complex groups could be compared with respect to their information-processing features. This manipulation was not expected to affect the risky shift, just as conceptual complexity is not correlated with risk conservatism (Choice Dilemmas scores).

Procedure

Discussion without consensus. The experimenter handed out booklets containing six of the Choice Dilemmas items and then gave instructions which were taken verbatim from Wallach and Kogan (1965, p. 10). The group was asked to discuss each choice situation for about 5 minutes. It was explained that the purpose of these discussions was to test the suitability of the items as material for a human-relations course, that is, whether they would bring out a sufficient variety of viewpoints. After the group had discussed the first item for about 5 minutes, the experimenter broke in at a natural break and asked them now to check their opinions privately. He emphasized that he was not interested in their prior decisions but in the opinion of each member now in view of the discussion. After everyone had noted his new response, they were asked to go on and discuss the next item, etc.

Discussion and consensus. After collecting the three booklets, the experimenter handed out another set containing the six remaining choice situations. The instructions, again taken verbatim from Wallach and Kogan (1965, p. 9), and again presented orally, asked the group now to arrive at a unanimous decision regarding the level of risk in each choice situation, and to put down a check indicating the group consensus for each item. After the group had gone through all six items this way, they were asked to provide new individual decisions, with instructions condoning the departure from the group decisions.

Alone Condition

Unnoticed by the group, the two subjects assigned to the alone condition were shown into two different rooms, each adjoining the discussion room. The subject in the *viewer* condition was placed behind a one-way mirror from where he could observe as well as listen to the group. The *listener* was placed in a room equipped with a loudspeaker transmitting the group discussion. Each of the subjects was given an instruction sheet and a booklet containing the 12 Choice Dilemmas items arranged in the order they were going to be taken up by the group. The instructions were exactly the same for the two conditions and read as follows.

> You will listen to a group discussing a set of problems which are described in the accompanying questionnaire and with which you are familiar from a previous testing session. You will first hear the instructions being read to the group and will then listen to the group discussing each of the choice situations. Each time the group has finished discussing an item, you are to indicate your own judgment—the minimum odds of success you would demand before recommending that the more attractive but riskier alternative be chosen.
>
> The purpose of this experiment is to find out whether a group discussion brings out new information—information which a listener may not have considered in making his initial judgments. This situation is similar to a panel discussion: Those attending it (the audience) expose themselves to a variety of points of view. The group discussion may or may not lead to a change—or a solidification—of your previous judgments. Make your decisions on the basis of your own thinking while taking cognizance of the opinions and arguments brought out in the group.

A few minutes after the experimenter had left the viewer and the listener in their respective rooms, he entered the discussion room together with the group and started giving instructions (see above), which were of course heard by the subjects in the alone condition.

RESULTS

The measure used in testing for the existence of a risky shift consists of the average, for a group, of the three initial overall risk scores, minus the corresponding average of the postdiscussion scores. The same method of analysis has been employed in Wallach and Kogan (1965). One-tailed, one-sample *t* tests were employed to determine the significance of the differences.

Table 1 shows that the groups shifted toward higher risk ($t = 2.73$, $p < .01$). The approximately equal magnitude of the shifts in the

TABLE 1

GROUP MEMBERS' SHIFTS TOWARD HIGHER RISK TAKING AS A RESULT OF DISCUSSION WITH OR WITHOUT CONSENSUS

	Prediscussion		Postdiscussion		
	$\bar{X}$[a]	*SD*	$\bar{X}$	*SD*	*t*
All 12 items (pre- vs. postdiscussion private decisions)					
All groups ($N = 16$)	66.82	4.52	59.74	6.80	2.73****
Low CC groups ($N = 8$)[b]	67.49	4.72	60.72	6.95	1.97***
High CC groups ($N = 8$)	66.16	4.54	58.76	6.98	1.79*
Discussion without consensus					
All groups	32.23	5.01	28.83	4.83	2.28***
Low CC groups	33.45	5.63	28.68	4.62	1.90***
High CC groups	31.01	4.33	28.99	5.43	1.12
Discussion with consensus (individual decisions vs. group consensus)					
All groups	34.59	4.50	30.19	4.97	2.53****
Low CC groups	34.04	5.71	31.12	5.87	1.38
High CC groups	35.15	3.20	29.25	4.13	2.01***
Discussion with consensus (individual decisions vs. postconsensus private decisions)					
All groups	34.59	4.50	30.91	4.75	2.39***
Low CC groups	34.04	5.71	32.04	4.83	1.30
High CC groups	35.15	3.20	29.77	4.70	1.92***
Discussion with consensus (group decision vs. postconsensus private decisions)					
All groups	30.19	4.97	30.91	4.75	−1.32
Low CC groups	31.12	5.87	32.04	4.83	− .87
High CC groups	29.25	4.13	29.77	4.70	−1.25

[a] High scores indicate low risk-taking.
[b] CC denotes conceptual complexity.
* $p < .07$.
*** $p < .05$.
**** $p < .01$.

TABLE 2

SHIFTS OF "VIEWERS" AND "LISTENERS" TOWARD HIGHER RISK TAKING

	Before discussion		After discussion		
	$\bar{X}$[a]	*SD*	$\bar{X}$	*SD*	*t*
All 12 items					
Viewers ($N = 16$)	67.25	6.33	59.94	6.51	2.80****
Listeners ($N = 14$)[b]	66.86	7.37	62.64	7.18	1.76**
Discussion without consensus					
Viewers	32.69	5.42	28.87	4.32	2.70****
Listeners	34.93	4.56	32.00	4.71	1.96***
Discussion with consensus					
Viewers	34.56	5.90	31.06	6.00	2.09***
Listeners	31.93	5.78	30.64	5.76	0.91

[a] High scores indicate low risk taking.
[b] Two subjects did not show up for the experiment, hence only 14 listeners.
** $p < .06$.
*** $p < .05$.
**** $p < .01$.

TABLE 3

COMPARISON OF SHIFTS IN THE GROUP AND ALONE CONDITIONS

	Group condition[c]		Alone condition[a]		
	$\bar{D}$[a]	*SD*	$\bar{D}$[a]	*SD*	*t*
All 12 items[b]	7.00	7.10	V 7.31	7.24	<1.00
			L 4.21	7.84	1.01
Discussion without consensus	3.13	4.74	V 3.81	3.76	<1.00
			L 2.93	4.70	<1.00
Discussion and consensus	4.63	5.19	V 3.50	1.28	<1.00
			L 1.28	5.09	1.74**
Discussion and consensus[b]	3.84	4.76	V 3.50	1.28	<1.00
			L 1.28	5.09	1.39

[a] $\bar{D}$ = pre- versus postdiscussion differences. Positive scores indicate increased risk taking after discussion.
[b] Difference scores for groups involve pre- versus postdiscussion private decisions.
[c] V = viewers ($N = 16$), L = listeners ($N = 14$). Group condition, $N = 16$.
** $p < .06$, 1-tailed test.

discussion-only and the discussion-and-consensus conditions—with a somewhat larger shift in the latter instance—confirm the findings from previous research (e.g., Wallach & Kogan, 1965), suggesting that a consensus requirement (a group decision) has a negligible effect on shift magnitude.

Table 2 presents evidence of a risky shift by the viewers ($t = 2.80$, $p < .01$), as well as the listeners ($t = 1.76$, $p < .06$). The listeners' shift on the discussion-and-consensus items is not significant.

Two-sample *t* tests were used in comparing the different conditions with regard to shift magnitude. Table 3 compares the group shifts with those of the viewers and the listeners. The shifts of group members and viewers are about equal in magnitude. The group shifts are somewhat more pronounced than those of the listeners. This trend is attributable to the shift difference on the items requiring a group consensus ($t = 1.74$, $p < .06$).

Table 4 presents *t* tests comparing the average shifts of the viewers with those of the listeners and shows that they do not differ significantly.

DISCUSSION

Physically separated individuals were found to shift to higher risk after merely witnessing a group's discussion of the given decision problems. This finding had been expected since the nonparticipant subjects were still privy to the information exchange, which, according to Brown, is the primary source of the risky shift. The Kogan-Wallach theory could explain the effect only if its terms were further broadened as follows. The subject behind the one-way mirror is in a position to receive all the affective cues emitted during the group session, enabling him to partake vicariously of the "emotionally tinged interpersonal connections," which, according to Wallach and Kogan (1965), are instrumental in bringing about the risky shift. Yet such an interpretation centers on the idea of affective bonds and bypasses the spread-of-responsibility notion. The question is then raised of whether the concept of responsibility is really applicable in view of the hypothetical nature of the Choice Dilemmas. For whom, to whom, in which way and to what extent are the group members responsible? Essential dimen-

TABLE 4

COMPARISON OF VIEWERS AND LISTENERS WITH REGARD TO SHIFT MAGNITUDE

	Viewers ($N = 16$)		Listeners ($N = 14$)		
	$\bar{D}$[a]	*SD*	$\bar{D}$	*SD*	*t*
All 12 items	7.31	7.24	4.21	7.84	1.10
Discussion without consensus	3.81	3.76	2.93	4.70	<1.00
Discussion with consensus	3.50	5.63	1.28	5.09	.87

[a] $\bar{D}$ = difference between post- and prediscussion scores. Positive scores indicate increased risk taking after witnessing the discussion.

sions of responsibility are absent in this decision context.

The inability of responsibility diffusion to account for the findings of this study does not preclude it as a cause of the shift effect. But these results support the conclusion that the diffusion of responsibility is not the primary or only source of the group risky shift.

REFERENCES

BROWN, R. *Social psychology.* New York: Free Press, 1965.

HINDS, W. C. Individual and group decisions in gambling situations. Unpublished master's thesis, Massachusetts Institute of Technology, School of Industrial Management, 1962.

KOGAN, N., & WALLACH, M. A. Risk taking as a function of the situation, the person, and the group. *New directions in psychology III.* New York: Holt, Rinehart & Winston, 1967.

NORDHØY, F. Group interaction in decision-making under risk. Unpublished master's thesis, Massachusetts Institute of Technology, School of Industrial Management, 1962.

RABOW, J., FOWLER, F. J., BRADFORD, D. L. HOFELLER, M. A., & SHIBUYA, Y. The role of social norms and leadership in risk taking. *Sociometry,* 1966, 29, 116–127.

SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. *Human information processing: Individuals and groups in complex social situations.* New York: Holt, Rinehart & Winston, 1967.

STONER, J. A. F. Comparison of individual and group decisions involving risk. Unpublished master's thesis, Massachusetts Institute of Technology, School of Industrial Management, 1961.

TEGER, A. I., & PRUITT, D. E. Components of group risk taking. Technical Report No. 18, 1966, Center of Research on Social Behavior, University of Delaware.

WALLACH, M. A., & KOGAN, N. Sex differences and judgment processes. *Journal of Personality,* 1959, 27, 555–564.

WALLACH, M. A., & KOGAN, N. The roles of information, discussion and consensus in group risk taking. *Journal of Experimental Social Psychology,* 1965, 1, 1–19.

WALLACH, M. A., KOGAN, N., & BEM, D. J. Group influences on individual risk taking. *Journal of Abnormal and Social Psychology,* 1962, 65, 75–86.

(Received July 25, 1966)

Journal of Personality and Social Psychology
1967, Vol. 6, No. 4, 472–476

ATHLETIC PARTICIPATION AND PERCEPTUAL AUGMENTATION AND REDUCTION

E. DEAN RYAN AND ROBERT FOSTER
University of California, Davis
Vanden High School, Travis Air Force Base

3 groups of high school boys, each group of 20 differing in athletic background, were tested for reaction time, pain tolerance, time estimation, and kinesthetic stimulation. There were significant differences between groups on all tests except reaction time and movement time, with the group composed of contact athletes (football and wrestling) tolerating most pain, underestimating time, and reducing subjective estimations of width on the kinesthetic test. The group of nonathletes (those who expressed a dislike for athletics) tolerated least pain, overestimated time, and had a tendency to enlarge the subjective judgment of width on the kinesthetic test. The noncontact athletes (tennis, golf, and track) fell between the other 2 groups on all tests. These results are in accord with Petrie's theory of augmentation and reduction.

Recently a significant relationship was found between the ability of an individual to tolerate pain and the type of athletic activity in which he chooses to participate (Ryan & Kovacic, 1966). Contact athletes (football players and wrestlers) tolerated more pain than noncontact athletes (tennis players and golfers), and the noncontact athletes in turn tolerated more pain than nonathletes. It was hypothesized that this relationship might have been due to differences in a general perceptual characteristic of "augmenting" or "reducing" sensory inputs. Petrie and associates have shown that certain individuals appear to consistently reduce the intensity of their perceptions, while other individuals tend to consistently augment the intensity of perception (Petrie, 1960). Augmentation and reduction have been estimated by measuring the change in kinesthetically perceived size after stimulation. The subjects match the width of a standard bar which is felt with one hand, with an area on a tapered bar which is felt with the other hand. A large (or smaller) block is then rubbed. Thereafter, the original test bar is again equated with the tapered bar, and the difference between the two subjective estimations is measured. Those individuals who tend to consistently reduce after stimulation have been shown to be more extroverted than those who augment (Eysenck, 1957), more tolerant of pain (Petrie, Collins, & Phillips, 1960), less tolerant of sensory deprivation (Petrie et al., 1960), more mesomorphic (Wertheimer, 1955), and to judge time as passing more slowly than the augmenters (Petrie et al., 1960). In addition, significantly more reducers were found in a delinquent group and significant fewer augmenters than would be expected by chance (Petrie, McCulloch, & Kazdin, 1962). Petrie suggested that these differences are partially due to the tendency to reduce or augment the perceptual intensity of stimulation in general. Any change in the intensity of stimulation may cause subsequent stimulation to appear subjectively less intense for the "reducer" and more intense for the "augmenters."

All of the characteristics of the reducer, the tolerance of pain, intolerance of sensory deprivation, mesomorphy, extroversion, and to an extent the characteristics of the delinquent, have frequently been associated with athletic groups. It may be that the subjective reduction of sensory stimulation is associated with the choice of activities a child engages in. If indeed, the reducers suffer from lack of stimulation as suggested by Petrie et al. (1962), then they would need change, movement, speed, and possibly body contact, rather than sedentary pursuits. Conversely, it may be that the type of activity an individual engages in influences his perceptual pattern, thus encouraging augmentation or reduction.

This study was designed to investigate the relationship between a general perceptual pattern of augmentation and reduction in groups differing in athletic interests. It was hypothesized that groups participaitng in contact sports would possess the perceptual pattern of the reducer and thus reduce most in their estimation of kinesthetically perceived size, have faster reaction times, faster movement times, judge time as passing slower, and would tolerate most pain. Groups not interested in athletics would possess the perceptual characteristics of the augmenter, and thus reduce less in their estimation of kinesthetically perceived size, have slower reaction times, slower movement times, judge time as passing faster, and tolerate less pain. Groups participating in noncontact sports would tend to fall between the other two groups on all tests.

METHOD

Subjects

A questionnaire was administered to male high school students. They were asked their likes, dislikes, hobbies, and recreational pursuits. On the basis of these answers three groups of subjects were selected. Group I was composed of boys who expressed a liking for certain athletics and were at the same time actually participating in contact sports (football or wrestling), Group II of boys who expressed an interest in noncontact sports and were also participating *only* in noncontact sports (golf, tennis, track), and Group III of boys who had expressed a dislike for athletics and were inactive as far as sports were concerned.

Letters were sent to parents requesting permission for sons to participate in an experiment dealing with perception. It was indicated that one of the tests would be a measure of pain tolerance. To obtain 20 subjects in each group it was necessary to contact 22 contact athletes, 26 noncontact athletes, and 26 nonathletes. The subjects assumed they had been randomly selected, and were unaware that athletic participation was a factor or that the experiment was in any way related to the questionnaire administered earlier.

Apparatus

Reaction times. The reaction-time apparatus consisted of a small base with a 16 inch upright. On the upright were mounted in a vertical position two small neon lights and a microswitch. A second microswitch was located in the base of the reaction timer. The upper light was amber in color and served as the ready signal. The experimenter could vary the ready period from ½ to 4 seconds. The onset of the lower green light, which served as the stimulus, activated two standard electric timers. The first timer was stopped as the subject's finger moved from the upper microswitch, and the second timer stopped as the microswitch mounted in the base was touched. Reaction time was read directly from the first clock; movement time was computed by subtracting the time of the first clock from that of the second.

Pain tolerance. An adaptation of Poser's mechanical stimulator was used (Poser, 1962). A single plastic, aluminum-tipped football cleat was secured to a curved fiber plate and fitted to the leg. The cleat was placed against the anterior border of the tibia, midway between the ankle and the knee. The sleeve of a standard clinical sphygmomanometer was used to secure the cleat firmly in place. Cleat pressure against the tibia was induced by inflating the armlet at a slow constant rate (approximately 5 millimeters of mercury per second) until the subject indicated verbally that he was no longer willing to endure the pain. Pressure was immediately released by deflating the cuff. The data were recorded in millimeters of mercury.

Augmentation and reduction. The apparatus used to measure augmentation and reduction was an adaptation of the kinesthetic aftereffect apparatus, originally described by Koehler and Dinnerstein (1947). The variable consisted of a wedge-shaped piece of hardwood, 30 inches long, 1 inch thick, 4.0 inches at the wide end, tapering to .5 inch at the narrow end. To maintain parallel alignment of the fingers, a sliding finger guide was fixed atop the wedge. Two hardwood blocks with parallel sides, 10 inches long and 2 inches thick, served as the standard bar and the stimulation bar. The standard bar was 1.5 inches wide and the stimulation bar 2.5 inches wide.

All subjects were tested individually and in a random order. The subject was told he would be given five different tests and was instructed to do as well as possible on each one. The sequence was the same for each subject, that is, reaction time, 2-minute time estimation, 20-second time estimation, pain tolerance, and the kinesthetic measures of augmentation and reduction.

At the conclusion of the reaction-time test the subject was asked to sit quietly with his hands resting in his lap. He was asked to judge a 2-minute period. When the subject thought 2 minutes had elapsed he would say, "time." This was repeated four times. No knowledge of results was given. The scores reported are the mean of the four trials.

Next, the subject was given a stopwatch and asked to duplicate the time interval demonstrated by the experimenter. The experimenter, holding the watch so the face was not visible to the subject, started his watch and let it run for 20 seconds. The subject, holding his own watch with the face down, was told to duplicate the time between the two clicks. This was done four times, each for a 20 second period, with the mean of all trials used for comparison.

At the completion of the 20-second test the subject was told his tolerance to pain would be measured, wherein he should make an effort to stand as much pain as possible. The pressure cleat was applied, and the subject was informed that pressure to the leg would be increased gradually and terminated when he could tolerate no more. The subject was unable to see the mercury column. After the first trial the experimenter commented that the score was considerably lower than the average of the group tested, and the subject was asked to take the test a second time, doing better if possible. The cleat was lowered 1 inch and a second trial was given.

Then the subject was blindfolded. He was instructed to grasp the standard bar between thumb and forefinger of the dominant hand and find the point that felt equal on the wedge-shaped comparison bar grasped identically in his other hand. Four base-line measurements were taken, alternating ascending and descending trials, that is, starting from the narrow end of the wedge on Trials 1 and 3 and from the wide end on the others. The standard bar was held in the subject's dominant hand during this testing period. The subject was then given the wider test block and instructed to rub with his preferred hand at a constant rate for 90 seconds. He then equated the perceived equivalent width of the original test block on the tapered bar. Again four measurements were taken, alternating ascending and descending trials, with the standard bar being held as in the original base-line measurements. This procedure was repeated for a second 90-second trial and finally for a 120-second trial.

RESULTS

The results are presented in Tables 1 and 2, with all means other than movement time being in the predicted direction. To test the significance of the difference between groups on each test a simple analysis of variance was used.

Augmentation and Reduction

There were no differences between groups in judging the width of the test bar prior to stimulation, with the contact athletes judging the 1.5 inch comparison bar as 1.67 inches, the noncontact as 1.69 inches, and the nonathletes as 1.66 inches. After 90 seconds of stimulation all three groups subjectively reduced the size of the comparison bar, although the difference between groups was not significant (Table 2). By 180 seconds of stimulation both athletic groups had increased the amount of reduction, while the trend for the nonathletic group had reversed, that is, they judged the comparison bar as being wider than after 90 seconds of stimulation. At this stage the difference between groups approached the .05 level of significance, $F\ (2{,}57) = 3.13, p < 3.15$. By 300 seconds of stimulation the difference between groups was quite pronounced, the estimation of width being .258 inch less for the contact athletes than before stimulation, for noncontract athletes .213 inch less, and for the nonathletes, .095 inch less. This difference was significant beyond the .01 level, $F\ (2{,}57) = 6.94$.

Reaction Time and Movement Time

Reaction times and movement times reported are the mean of all 20 trials. While the mean reaction times for the three groups were in the predicted direction, with the contact athletes fastest, noncontact athletes next, and the nonathletes slowest, the difference between the three groups was not significant, $F\ (2{,}57) = 1.61$ (Table 1). For movement time no trend was apparent, and mean scores were so similar that no test of significance was necessary.

Time Estimation

The differences between groups in estimation of 120 seconds was highly significant, $F\ (2{,}57) = 19.19$, with the contact athletes reducing the period by 16.44 seconds, the noncontact athletes underestimating by only 0.91 second, and the nonathletes enlarging by 2.26 seconds (Table 1). When duplicating a 20-second period the trend was the same, with the contact athletes reducing 1.69, the noncontact athletes enlarging 0.73 second, and the nonathletes enlarging 1.50 seconds. The difference between groups was highly significant, $F\ (2{,}57) = 12.86$ (Table 1).

Pain Tolerance

On Test 1 the amount of pressure the three groups were willing to tolerate was in the predicted direction; the difference between groups

TABLE 1

TIME MEASURES FOR ATHLETIC SUBGROUPS

Test	Contact athletes	Noncontact athletes	Nonathletes	F
Mean RT (sec.)	.201	.208	.213	1.61
Mean MT (sec.)	.117	.123	.117	—
Mean estimates of 120 sec.				
Trial 1	96.55	111.50	114.90	
Trial 2	101.95	119.35	122.50	
Trial 3	110.85	119.90	125.00	
Trial 4	104.90	125.20	126.85	
Mean of 4 trials	103.56	119.09	122.26	19.19**
Mean estimates of 20 sec.				
Trial 1	18.05	21.15	21.75	
Trial 2	19.30	20.10	21.65	
Trial 3	18.50	21.15	21.20	
Trial 4	17.40	20.50	21.40	
Mean of 4 trials	18.31	20.73	21.50	12.86**

$^{**}p < .01$.

TABLE 2

DIFFERENCES IN KINESTHETIC SIZE PERCEPTION AND IN PAIN TOLERANCE FOR ATHLETIC SUBGROUPS

	Contact athletes	Noncontact athletes	Nonathletes	*F*
Mean change in kinesthetically perceived size after stimulation				
90-sec stimulation	−.110 in.	−.108 in.	−.103 in.	*ns*
180-sec. stimulation	−.168	−.173	−.065	3.13*
300-sec. stimulation	−.258	−.213	−.095	6.94**
Mean pain tolerance (mm/Hg)				
Trial 1	286.75	231.00	207.50	14.27**
Trial 2	298.00	251.00	230.75	13.78**

* $p < .06$.
** $p < .01$.

being highly significant, $F\ (2,57) = 14.27$. The group of contact athletes tolerated 285.75 millimeters of mercury, the noncontact athletic group 231.00, and the nonathletic group, 207.50. On the second trial, after being individually informed they had done poorly, all three groups improved, with the contact athletic group averaging 298.00 millimeters, the noncontact athletic group 251.00, and the nonathletic group, 230.75. The difference between groups was highly significant, $F\ (2,57) = 13.78$ (Table 2). It was impossible to make a meaningful comparison of the amount of change each group made from Trial 1 to Trial 2 as the mercury column on the sphygmomanometer only read to 300 millimeters. It should be noted that the mean score on Trial 2 for the contact athletes was 298 millimeters of mercury. Of the 20 contact athletes tested, 16 withstood the maximum pressure, while each of the remaining 4 subjects tolerated 290 millimeters. Of the noncontact athletes five tolerated the maximum pressure, while three nonathletes tolerated the maximum.

Comparison of Extreme Augmenters and Reducers

Due to the large individual differences apparent within each group, and because Petrie's work has dealt primarily with extreme augmenters and reducers, an analysis was made of the 15 subjects making the least reduction in their estimation of kinesthetically perceived size after 300 seconds of stimulation and the 15 subjects making the greatest reduction, regardless of their athletic background. The results shown in Table 3 parallel and substantiate the work of Petrie. There was no difference between groups in reaction time and movement time, but all other differences were in the predicted direction and highly significant. Of the extreme reducers, 7 were contact athletes, 6 noncontact athletes, and 2 nonathletes. Of the extreme augmenters, 11 were nonathletes, 1 noncontact athlete, and 3 contact athletes.

DISCUSSION

The results of this study clearly support Petrie's theory of a generalized tendency for certain individuals to consistently reduce or diminish their perception of stimulation and for others to consistently augment or enlarge perceptions. The contact athletes showed the characteristics of the reducer, making the greatest subjective reduction of kinesthetically perceived size after stimulation, tolerating most pain, and consistently judging time as passing more slowly, than did groups composed of noncontact athletes or nonathletes.

While differences in pain tolerance could be explained by simply assuming athletes were more motivated to withstand pain, time estimation and estimation of kinesthetically perceived size are less amenable to changes in motivation. No amount of conscious effort should induce a naïve subject to vary time or kinesthetic sensitivity in one direction or another. One other point should be noted. During the testing period the experimenter was aware of the subject's classification.

TABLE 3

COMPARISON OF EXTREME AUGMENTERS AND REDUCERS AFTER 300 SECONDS OF STIMULATION[a]

Test	Reduced most	Reduced least	*t*
Kinesthetic estimation	−.380	+.018	
RT	.21 sec.	.21 sec.	0.45
MT	.12 sec.	.12 sec.	0.97
Time estimated 20 sec.	19.48 sec.	21.72 sec.	3.60**
Time estimated 120 sec.	105.96 sec.	126.66 sec.	4.30**
Pain tolerance 1	267.64 mm/Hg	223.98 mm/Hg	3.55**
Pain tolerance 2	280.31 mm/Hg	245.31 mm/Hg	3.41**

[a] $N = 15$ for both groups.
** $p < .01$.

Since procedures were standardized for each test it is unlikely that the results could be attributed to experimenter bias. The point remains, however, that this must be considered as a possible source of error.

The usual response in estimating kinesthetically preceived size is for subjects to reduce the perceived size of a test block after stimulation with a larger block, and to enlarge the perceived size of a test block after stimulation by a smaller block. Petrie has suggested, however, that

> the reducer will tend to reduce even if he is stimulated for equal amounts of time with a block that is larger than the test object and a block that is smaller than the test object; that is to say, the total effect is for him to reduce [Petrie, Holland, & Wolk, 1963].

The opposite relation holds for the augmenter, who will regularly tend to enlarge his estimate of size. In this study only large block stimulation was used, thus minimizing the chances of finding augmentation. In spite of this, however, 30% of the nonathletic group augmented in the kinesthetic estimation of size. It should be noted further that the nonathletic group, or the augmenters, consistently overestimated time.

Two inconsistencies in experimental method employed by Petrie and by this study should be noted. First, studies by Petrie have employed only ascending trials on the tapered bar, that is, the subject always started from the smaller end of the bar. In the present study, alternating ascending and descending trials were used. Dinnerstein et al. (1962), studying groups that differed in pain tolerance, found no difference in figural aftereffects when using alternating ascending and descending trials, but found significant differences when only ascending trials were analyzed. Thus it would be expected that even greater differences would be found between athletic subgroups if methods used by Petrie had been followed. Second, Petrie suggests a minimum of 45 minutes rest before the kinesthetic test is administered. "Such a resting period is essential in order to allow the wearing off of the effect of whatever the subject may have been handling prior to testing [Petrie et al., 1962]." In the present study approximately 30 minutes elapsed from the end of the reaction-time experiment to the start of the kinesthetic test. For approximately 5 minutes of this period the subject held a stopwatch in his preferred hand, pressing with his thumb four times. Whether the stimulation of the reaction-time test, followed by only 30 minutes of rest, had an effect on kinesthetic sensitivity is problematical. However, in light of the fact that there were no differences between groups in judgment of width prior to stimulation it would appear that differences due to time would be negligible.

The similarities between the perceptual characteristics of the contact athlete and the juvenile delinquent should be noted, and as Petrie points out, education of these individuals needs to make allowance for their vulnerabilities and strengths. Both groups appear to need stimulation, movement, and change instead of more sedentary activities. It would be expected that confinement and restriction of movement in the typical classroom would be less conducive to good study habits for the reducer than for the augmenter. Thus, an entirely different pedagogical approach would be desirable for the opposing perceptual types. Further, it would seem that a program of vigorous activity might provide a socially acceptable means of relieving sensory monotony and thus reduce the need for less desirable stimulation.

REFERENCES

DINNERSTEIN, A. J., LOWENTHAL, M., MARION, R. B., & OLIVO, J. Pain tolerance and kinesthetic aftereffect. *Perceptual and Motor Skills,* 1962, **15**, 247–250.

EYSENCK, H. J. *The dynamics of anxiety and hysteria.* London: Routledge & Kegan Paul, 1957.

KOEHLER, W., & DINNERSTEIN, D. Figural aftereffects in kinesthesis. In A. Michotte (Ed.), *Miscellanea Psychologica.* Paris: Libraire Philosophique, 1947. Pp. 196–220.

PETRIE, A. Some psychological aspects of pain and the relief of suffering. *Annals of the New York Academy of Science,* 1960, **86**, 13–27.

PETRIE, A., COLLINS, W., & SOLOMON, P. The tolerance for pain and sensory deprivation. *American Journal of Psychology,* 1960, **73**, 80–90.

PETRIE, A., HOLLAND, T., & WOLK, I. Sensory stimulation causing subdued experience: Audio-analgesia and perceptual augmentation and reduction. *Journal of Nervous and Mental Disorders,* 1963, **137**, 312–321.

PETRIE, A., MCCULLOCH, R., & KAZDIN, P. The perceptual characteristics of juvenile delinquents. *Journal of Nervous and Mental Disorders,* 1962, **134**, 415–421.

POSER, E. G. A simple and reliable apparatus for the measurement of pain. *American Journal of Psychology,* 1962, **75**, 304–305.

RYAN, E. D., & KOVACIC, C. R. Pain tolerance and athletic participation. *Perceptual Motor Skills,* 1966, **22**, 383–390.

WERTHEIMER, M. Figural aftereffect as a measure of metabolic deficiency. *Journal of Personality,* 1955, **24**, 56–73.

(Received July 30, 1966)

Diary No 96
Date......10...9...70
Lib
Bureau Ednl. Psy. Research

Journal of Personality and Social Psychology

Volume 6, No. 4
Part 2 of 2 Parts

August, 1967
Whole No. 638

MONOGRAPH

EFFECTANCE AROUSAL AND ATTRACTION [1]

DONN BYRNE AND GERALD L. CLORE, JR.[2]

University of Texas

The motivational construct proposed by both cognitive and reinforcement theorists to account for the relationship between attitude similarity-dissimilarity and attraction is tentatively equated with White's effectance motive. In a series of investigations utilizing over 600 Ss, there was an attempt to test some of the implications of this formulation. 2 different experimental approaches to effectance arousal were employed, and a scale was devised as an index of arousal. Among other findings, a moderate level of effectance arousal resulted in an increase in the slope of the attitude-attraction function, while a high arousal level decreased the slope of the function. In addition, the arousal properties of attitude statements were confirmed.

In the experimental investigation of interpersonal attraction, a strikingly consistent antecedent-consequent relationship has been established: attraction toward a stranger varies as a linear function of the proportion of that stranger's attitudes which are congruent with those of the subject (e.g., Byrne, 1962; Byrne & Clore, 1966; Byrne & Griffitt, 1966; Byrne & Nelson, 1965; Byrne & Rhamey, 1965).

Such empirical relationships are, of course, open to a variety of theoretical interpretations. In this instance, the proposed formulations may be roughly categorized as adhering either to a cognitive model (e.g., Heider, 1958; Newcomb, 1959) or to a reinforcement model (e.g., Clore, 1966; Golightly & Byrne, 1964; Nelson, 1965). Despite numerous differences in these two approaches, there is one aspect of the similarity-attraction relationship about which proponents of the two types of theorizing appear to be in agreement. The effect of attitudinal material on attraction is attributed to the mediation of motivational factors. In the present paper, the hypothesized motive system will be examined in some detail along with a series of experiments designed to provide empirical substance for such a formulation.

ATTITUDES AND MOTIVE AROUSAL

The two major cognitive explanations of attraction involve tridimensional systems consisting of the subject, another individual, and an object about which they each hold attitudes or beliefs. Newcomb designated the three components as A-B-X, while Heider used a *p-o-x* notation. Newcomb (1953, 1959, 1961) and Heider (1958) each proposed that when there is a positive relationship between two individuals, the discovery of attitudinal

---

[1] This research was supported in part by Research Grant MH-11178-02 from the National Institute of Mental Health, United States Public Health Service. The assistance of the following individuals is gratefully acknowledged: Dale Baskett, William Griffitt, Carol Hampton, Mark Schwartz, Robert Shaw, and Daniel Stefaniak.

[2] Now at the University of Illinois.

Copyright © 1967 by the American Psychological Association, Inc.

discrepancy is "disturbing" or "threatening" and is a "danger signal." The disturbed state is described by the construct *strain* which ". . . refers to a hypothetical state of psychological tension occurring under certain conditions of judged discrepancy between own and another's attitude toward the same object [Newcomb, 1959, p. 393]." Strain or imbalance serves as an instigation to behavior which serves to restore a balanced state. Among the possible alternative responses (e.g., increased communication, misperception, attitude change) is a reduction in the individual's attraction toward the other person.

As a result of the Texas studies (beginning with Byrne, 1961), it has become possible to place the relationship between attitude similarity and attraction in an antecedent-consequent framework by experimentally manipulating the attitudes of a stimulus person and determining the subsequent effects on attraction toward him. Among other considerations, it is now clear that a previously established positive relationship between A and B (or *p* and *o*) is *not* a prerequisite for the observed effects. Rather, similar and dissimilar statements from a complete stranger elicit attraction responses in a manner consistent with that found in the A-B-X designs. Thus, attitude statements and attraction may be conceptualized in a stimulus-response framework with a motivational state as a hypothetical intervening construct (Byrne, 1966).

What are the implications of including a motivational construct in this S-R relationship? On the one hand, an intervening variable can simply represent a higher-order generalization which permits one to conceptualize and organize a variety of quite different specific stimuli and specific responses within a single unifying framework. In addition, a motivational construct suggests that one should be able (*a*) to specify the stimulus conditions under which drive level is increased or decreased, and (*b*) to demonstrate specific drive stimulus properties for this motive (i.e., subjects should make different responses under the indicated stimulus conditions from those made under other types of arousal conditions). Before proceding to such evidence, it would be helpful to examine more specifically the hypothesized characteristics of the drive in question.

EFFECTANCE MOTIVE

It is proposed that the motive which is presumably activated by attitudinal material is not unique to that particular stimulus situation. Rather, the motivational properties of attitude statements represent a special instance of a more general phenomenon. At the risk of erring by "the inclusion in the same category of events which only superficially belong together . . . [Dember & Earl, 1957, p. 91]," the authors suggest that the type of secondary motive described in the attraction studies has appeared under various labels and in various contexts. Examples are the need for cognition or the need to experience an integrated and meaningful world (Cohen, Stotland, & Wolfe, 1955), the need to be able to know and predict the environment (Kelly, 1955; Pervin, 1963), the desire for certainty which involves understanding the environment and making it predictable (Brim & Hoff, 1957), and the drive to evaluate one's own opinions and abilities (Festinger, 1954). It has previously been suggested (Byrne, Nelson, & Reeves, 1966) that these various concepts all fit within White's (1959) *effectance motive* in that each involves a process related to effective interaction with the environment. White (1959) said, for example:

> . . . it is maintained that competence cannot be fully acquired simply through behavior instigated by [primary] drives. It receives substantial contributions from activities which, though playful and exploratory in character, at the same time show direction, selectivity, and persistence in interacting with the environment. Such activities in the ultimate service of competence must therefore be conceived to be motivated in their own right. It is proposed to designate this motivation by the term effectance, and to characterize the experience produced as a *feeling of efficacy* [p. 329].

It should be noted that there is a major difference between White's emphasis and that suggested here. That is, White stressed the "positive" aspects of the motive to explain why organisms avoid the monotonous and repetitious and familiar in order to seek stimulation via exploration, play, intellectual curiosity, and manipulation of the environment (see, for example, Jones, Wilkinson, & Braden, 1961). The present authors argue, however, that the same motivational construct which accounts for a preference for stimulation also accounts for a negative response to stimuli which lie further along the continuum of unfamiliarity, unpredictability, and unexpectedness. The tendency of organisms to prefer stimuli which lie at an optimal point between familiar-simple and unfamiliar-complex has been described by a number of theorists (e.g., Hebb, 1949; Leuba, 1955; McReynolds, 1956; Munsinger & Kessen, 1964; Woodworth, 1958).

Acquiring the Effectance Motive

A speculative description of the development of such a motive system has been provided by Dollard and Miller (1950) in discussing the learned drive to be logical and to make a correct report of the environment.

> While learning to talk the child receives training in matching words, sentences, and descriptive paragraphs with important features of the environment. He is trained not to call the sky "green" or the grass "blue." He is corrected for mistakes, and eventually learns to have anxiety about responding differently from other people. He is also taught to distinguish between the make-believe and the real—in other words, between the situations in which sentences that do not run parallel to sequences in the environment will be tolerated and those in which they will not. He is taught to discriminate between his own private dreams, images, and phantasies, and the perceptions that can be verified by cross-checking with other senses and that agree with the reports of other people . . .
>
> We believe that this training serves to provide motivation against hallucinations and delusions. The strength of this motivation is shown by the anxiety that a normal person will show if through some unexpected illusion he seems to be seeing things others do not see [p. 119].

At the anecdotal level, one might note that small children appear to experience pleasure in simply learning the correct verbal label to apply to any new object or event, in repeating endlessly a previously experienced sequential activity such as a game or a familiar story, or in grasping the meaning of that which was previously puzzling. In a similar vein, Woodworth (1958) described such behavior and suggested its utility:

> Its long-range value as the means of making the child acquainted with the world he has to deal with later, and so equipping him through play for the serious business of life, can scarcely lie within the little child's horizon. His goals are more limited and direct: to see this or that object more closely, to find what is behind an obstacle, to hear the noises an object makes when it strikes the floor, to be told the name of a thing or person [p. 78].

If, however, the stimulus events are sufficiently unfamiliar, unpredictable, or inexplicable, they arouse anxiety and fear. The desire to deal effectively with the environment can thus lead either to pleasurable exploration or to fearful withdrawal.

Adulthood does not alter these relationships. Rather, the well-adjusted adult simply has a wider sphere of knowledge and familiarity than does the child so that the physical world offers fewer day to day surprises which either challenge or frighten him. As White (1959) indicated,

> The building of houses, roads and bridges, the making of tools and instruments, the domestication of plants and animals, all qualify as planful changes made in the environment so that it comes more or less under control and serves our purposes rather than intruding upon them [p. 324].

In addition, in the terminology of Munsinger and Kessen (1964), the adult has developed relatively effective coding rules which increase the amount of environmental variation which can be processed. Nevertheless, when adults are confronted with unfamiliar, unpredictable, or inexplicable events, it appears that the consequent distress equals in intensity that of childhood.

Social Reality and the Effectance Motive

Thus, well-adjusted adults in a relatively stable, civilized society should be protected against unpleasant levels of effectance arousal by effective social training in childhood, years of experience in which to build up coding capacity, and a stable environment with built-in predictability. There may be occasional minor jolts (momentary forgetting of the location of one's car after a football game, minor loss of orientation just after moving to a new house, seeing an unidentified flying object) and at times major disruptions (economic depression, earthquake, war) which prevent in varying degrees the satisfaction of the effectance motive. For most of us most of the time, physical reality does not actually pose major problems related to effectance. Nevertheless, there remains what Festinger (1950) called "social reality," in which there is a much greater degree of unpredictability, lack of necessary knowledge, and the occurrence of seemingly illogical events. Dollard and Miller (1950) noted:

> The training in matching words to important features of the environment applies to the social as well as the physical environment. The child is trained to give an accurate account of social events and to make his statements about the probable social consequences of acts match the cultural pattern [p. 119].

Included here are all of the major aspects of our world about which we hold attitudes, opinions, beliefs, and values. Is there a God? What are the consequences of assuming an aggressive military posture with respect to Communist China? Is it reasonable to enjoy television comedies? Has the Civil Rights movement gone too far or not far enough? How should one raise his children? Is op art just a passing fad? To what should the sexual practices of adolescents be limited? It is obvious that the list can be extended almost endlessly and that the resolution of such questions in terms of discovering the correct answer, mastering the problem, dealing effectively with the issue, or testing the empirical consequents of one's assumptions is not possible. Thus, the raising of such issues should arouse the effectance motive. Further, the only relevant evidence to bring about effectance reduction lies in consensual validation or invalidation. The former occurs when others agree with or share our outlook and hence provide satisfaction of the effectance motive. Consensual invalidation occurs when others disagree with us and hence frustrate the satisfaction of the effectance motive.

It is suggested, then, that the attitudes of others are important to us because human beings depend on one another as sources of information about countless aspects of the environment. People are rewarding to one another as ". . . suppliers of new information and as confirmers or correctors of old . . . [Newcomb, 1959, p. 386]." Heider (1958) stated, "The power of similarity of beliefs or attitudes derives from the identity of the environment to which they refer, and from the fact that it is satisfying to find support for one's own view [p. 196]." Similarly, Festinger (1950) proposed, "If there are other people around him who believe the same thing, then his opinion is, to him, valid. If there are not others who believe the same thing, then his opinion is, in the same sense, not valid [p. 272]."

It might be noted that there is one sort of empirical support for the motivational interpretation of attitude material and for its possible equivalence to quite different stimuli. Similar attitude statements are found to act as positive reinforcers and dissimilar attitude statements to act as negative reinforcers in a discrimination-learning task (Byrne, Young, & Griffitt, 1966; Clore, 1966; Golightly & Byrne, 1964). In a similar way, predictable stimuli are found to be preferred to nonpredictable stimuli even though predictability is not instrumental in altering the situation (D'Amato & Gumenik, 1960; Lanzetta & Driscoll, 1966; Pervin, 1963). More direct evidence is, of course, needed.

In terms of providing a descriptive and speculative rationale for the effects of attitude similarity-dissimilarity on attraction, the notion of effectance may or may not provide instant *Verstehen.* The research problem, however, is to seek a means to arouse this hypothesized motive independently of the attitude situation, to devise a reliable behavioral index of arousal, and to determine experimentally the way in which effectance arousal influences the attitude-attraction relationship. The initial steps in such a research program are described in the remainder of this paper.

I. EFFECTANCE AROUSAL: STIMULI AND RESPONSES

If it can tentatively be accepted that a single construct labeled effectance encompasses the diverse situations outlined in the preceding discussion, it would follow that any situation which provides evidence of one's predictive accuracy, ability to understand, correctness, logicality, reality orientation, behavioral appropriateness—any information which permits or indicates effective functioning—would satisfy the effectance motive. Any situation providing the opposite type of evidence would frustrate the effectance motive. More specifically, with human and infrahuman subjects, investigators have defined (nonattitudinal) arousal conditions in terms of stimuli which are unfamiliar or strange (Berlyne, 1955; Dollard & Miller, 1950; Welker, 1956; Zimbardo & Miller, 1958), ambiguous (Brim & Hoff, 1957; Cohen et al., 1955), variable (Munsinger & Kessen, 1964), uncertain (Gerard, 1963; Lanzetta & Driscoll, 1966), unexpected (Maddi, 1961), unpredictable (Elliott, 1966; Gergen & Jones, 1963), or a discrepancy from an expected outcome (Carlsmith & Aronson, 1963; Dember & Earl, 1957).

One approach to the direct measurement of arousal is the utilization of the subject's verbal report. What content would be crucial as an index of arousal? Any situation in which the effectance motive is aroused should be unpleasurable and should evoke a negative emotional response. In addition, specific cognitive elements over and above an unpleasant feeling might be expected to accompany arousal. On the basis of previous research and theory, five somewhat distinct characteristics have been suggested.

1. *Teneness or uneasiness*. Stimuli which arouse effectance should evoke a general tension (Cohen et al., 1955) or uneasiness (Gergen & Jones, 1963; Welker, 1956). Dollard and Miller (1950) used the word "uncomfortable."

2. *Confusion*. Terms such as "uncertainty [Munsinger & Kessen, 1964]" and "desire for certainty [Brim & Hoff, 1957]" suggest a state of confusion about the interpretation of stimuli.

3. *Sense of unreality*. The effects of arousal might be expected to resemble the primitive feelings of childhood, including a parataxic or autistic sense of unreality. Not only should high effectance level be associated with childhood, but an aroused state should be more characteristic of schizophrenics than normals. "When training in the techniques and learned drives involved in clear thinking is defective, the individual is often described as having a weak Ego or a poor sense of reality [Dollard & Miller, 1950, p. 122]."

4. *Dream-like feeling*. Because private cue-producing responses are not directly observed by others and hence are less subject to social training, dreams ". . . are (as Freud discovered) less inhibited than waking thought and also are often illogical and unintelligible by ordinary standards of narration [Dollard & Miller, 1950, p. 123]." Thus, effectance arousal should be perceived as resembling the dream state.

5. *Desire for social comparison*. Possibly as a secondary consequence of the arousal state, it would seem that Festinger's (1950) social comparison process would also be operative in that individuals would desire to communicate with and compare themselves with others.

It was hypothesized, therefore, that the effectance motive is aroused by stimulus conditions which are unpredictable (and hence novel, unfamiliar, ambiguous, and unexpected), and that arousal is associated with verbal reports of uneasiness, confusion, unreality, dream-like feelings, and a desire for social comparison.

Procedure

In order to have a standard, controlled stimulus with nonpredictable qualities, the senior author and two National Science Foundation undergraduate research fellows[3] devised an 8-millimeter sound color movie. The rationale was simply to produce a series of scenes having no meaningful interrelationship and in which the sequence of events within a scene followed no logical schema. The three investigators worked independently, and the resultant edited version consists of 10 minutes and 23 seconds of continuing meaninglessness. Scenes include ceramic figures of cannibals cooking a missionary, an aerial view of a toy battleship, Mr. Ed the talking horse, a flushing toilet, a chess game played with cosmetic bottles, Negro children playing, a girl swimming, a dizzying ride through tree-tops, and a variety of animals. The sound track is predominantly that of *Voodoo Suite* played by Perez Prado with occasional interpolations of other sounds. This movie will be designated as Nonpredictable Movie I.

As an index of motive arousal, five 5-point self-rating scales were built in an effort to tap some of the hypothesized subjective elements of effectance arousal. These items dealt with feelings of unreality, feelings like those when dreaming, uneasiness, confusion, and the desire to know the thoughts of others. Each item is scored 1–5 and the total effectance arousal score can range from 5 to 25. The five effectance items plus four buffers were presented immediately after the movie was shown in a mimeographed form entitled, "Movie Reaction Scale" (see Appendix A).

---

[3] This portion of the research was supported by a grant from the National Science Foundation (G-E 1894) for undergraduate science education under the direction of Fillmore Sanford. The research fellows were Carol Hampton and Daniel Stefaniak.

Two other movies were employed for purposes of control. The neutral predictable movie was a documentary sound color 16-millimeter film entitled *Life in Morocco* (Pat Dowling Company). This is an 11-minute documentary which depicts rural and urban life in Morocco, accompanied by appropriate music and narration. Because the nonpredictable movie was hypothesized to have quite specific arousal properties, it was also necessary to compare its effects with those of a predictable film which nevertheless had drive-arousing properties. The film selected for this purpose was a medical film entitled, *Extracapsular Cataract Operation with Peripheral Iridectomy* (Visual Education Service, University of Minnesota) which depicts the gruesomely detailed procedures of a cataract operation in full color.

The subjects consisted of 120 male and female introductory psychology students at the University of Texas. They were assigned randomly to one of the three movie groups and run in small groups of 10–15 subjects. In the experimental sessions, the following instructions were given:

> There is an increasing interest in movies and television for instructional purposes. We are interested in what kind of learning takes place while a person watches a film. In a moment you will see a movie, and after it is over you will be asked about your understanding of it. Since we are interested in the learning process, it is important that you do not talk to each other or share your reactions either during the movie or after it is over.

Results

The means and standard deviations of the effectance arousal score for the three movies are shown in Table 1. A simple one-way analysis of variance indicated highly significant differences in arousal ($F = 38.12$, $df = 2/117$, $p < .001$). Further, the nonpredictable movie group was significantly more aroused than the group seeing the eyeball operation ($t = 4.11$, $df = 78$, $p < .001$). It seems reasonable to suggest that the greater arousal in the latter group than in the group viewing the neutral predictable movie ($t = 4.71$, $df = 78$, $p < .001$) is attributable to the somewhat unfamiliar and unpredictable aspects of a cataract operation for these subjects. It also might be noted that on two of the buffer items (disgusted and anxious), the subjects who viewed the two arousing movies did not differ, but both were significantly higher than the group viewing the neutral predictable movie.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF EFFECTANCE AROUSAL SCORES IN RESPONSE TO EXPERIMENTAL AND CONTROL MOVIES

	Neutral predictable movie	Arousing predictable movie	Nonpredictable movie
M	7.20	9.60	12.48
SD	1.49	2.79	3.38

Subjects were also asked to give their subjective response to the movies by indicating the one word which best described their feelings and also by describing their interpretation of the film. The most frequently given words to the nonpredictable movie were "confused," "interest," "curiosity," "bewilderment," and "strange," with a large array of other words given at least once including "reality," "disturbed," "insecurity," "anxiety," "tension," "nauseated," and "mixed-up." Interpretations of the meaning of the nonpredictable movie included themes of the passage of time, variety and oddity in everyday life, accelerated pace of living, difference between savagery and civilization, what is around one but not usually seen, and the decision by three subjects that there was no point at all. The interpretations of the two control movies were straightforward. For example, 27 of the 40 subjects viewing *Life in Morocco* said the one word best describing their feelings was "interested."

The internal consistency of the effectance arousal score was also investigated, utilizing some of the present subjects plus additional ones from subsequent samples. There were 120 subjects who had seen Nonpredictable Movie I, 80 who had seen Nonpredictable Movie II (to be described shortly), and 120 who had seen the neutral predictable movie. Each of the five items was evaluated in terms of its correlation with the total arousal score on the *other four items* combined. The correlations were .42 ($p < .001$) for Unreality, .22 ($p < .001$) for Dreaming, .38 ($p < .001$) for Uneasy, .55 ($p < .001$) for Confused, and .31 ($p < .001$) for Other's Thoughts. The corrected split-half (odd-even) reliability was found to be .69.

Discussion

While the nonpredictable movie had the hypothesized effects on the subjects' self-ratings, the relationship between these particular operations and the theoretical explanations of the attraction data remains to be demonstrated.

In the attitude studies, the negative effect of disagreement is attributed to the arousal and nonsatisfaction of the effectance motive. Schwartz (1966) reasoned that if the nonpredictable movie arouses effectance, anyone identified with the arousal stimulus would elicit negative attraction responses. In his doctoral research, Schwartz presented Nonpredictable Movie I to high school students, describing the movie as the product of a graduate student in a special film-making class. In a pilot study, it was found that as effectance arousal increased, attraction toward the supposed creator of the film decreased. In his primary experiment, Schwartz replicated this finding and also found significantly greater attraction toward the individual to whom a predictable control film was attributed than toward the producer of the nonpredictable film. Thus, the hypothesized effect of a nonpredictable stimulus on attraction was confirmed. The next research step was to determine whether exposure to this nonpredictable stimulus prior to participation in an attitude experiment would influence attraction.

II. THE INFLUENCE OF A NONPREDICTABLE MOVIE ON RESPONSE TO ATTITUDE SIMILARITY-DISSIMILARITY

It is quite conceivable that the "effectance arousal" score represents simply an index of relatively unpleasant affect. Thus, subjects respond negatively to the unpredictable movie and to whoever made the film. If this explanation is correct, exposure to the unpredictable movie just prior to examining the attitude responses of a stranger should simply have a negative effect on attraction responses. There might still be a linear relationship between similarity and attraction, but with an overall displacement in the negative direction.

If, on the other hand, the effectance interpretation is the accurate one, a different prediction would be made. In two previous studies, it has been found that the slope of the similarity-attraction function increases as the presumed effectance-arousing qualities of the attitudinal material increases (Byrne et al., 1966; Byrne & Rhamey, 1965). In a similar way, effectance motive as aroused by the nonpredictable movie would be expected to interact with effectance motive as aroused by the attitudinal material. Therefore, it was hypothesized that, compared to subjects in a neutral condition, subjects in an effectance-arousal condition respond more positively to a stranger with similar attitudes and more negatively to a stranger with dissimilar attitudes.

Method

The subjects were 80 introductory psychology students (40 males, 40 females) at the University of Texas. Early in the semester these students had taken an attitude survey consisting of 12 6-point items dealing with a variety of issues such as political parties, birth control, and careers for women. Each subject was randomly assigned to one of two effectance-arousal conditions and to one of two attitude-similarity conditions.

In small groups, subjects received the movie instructions as presented in the first investigation and then were shown either Nonpredictable Movie I or *Life in Morocco*. Immediately afterward, an "interpersonal judgment" experiment was conducted, purportedly as an interpolation between the learning experience (the movie) and questions to be asked about it afterward. The subjects were told that they would be given an attitude scale (name removed) filled out by another student, and their task was to form a series of judgments about this person. After studying the 12 responses made by the stranger, they were to rate him or her on the Interpersonal Judgment Scale (Byrne, 1966). This instrument consists of six 7-point scales dealing with the stranger's intelligence, knowledge of current events, morality, and adjustment and with the degree to which the subject would probably like this person and enjoy working with him. The latter two items constitute the measure of attraction; they are summed to yield an attraction score ranging from 2 to 14 with a split-half reliability of .85 (Byrne & Nelson, 1965). The "strangers'" scales were actually prepared by the experimenters such that each subject responded to a stranger who was either similar to himself on 10 items and dissimilar on 2 or similar on 2 and dissimilar on 10.

After judging the stranger, subjects were given the Movie Reaction Scale which contained the five-item measure of effectance arousal.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF ATTRACTION SCORES OF HIGH- AND LOW-AROUSAL SUBJECTS TOWARD SIMILAR AND DISSIMILAR STRANGERS

Effectance-Arousal level	Proportion of similar attitudes					
	.17		.83		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
High	7.67	1.97	9.88	1.76	8.55	2.18
Low	6.00	2.69	11.33	2.06	9.20	3.50
Total	7.00	2.43	10.75	2.07		

Results

The first analysis of the data consisted of a 2 × 2 factorial analysis of variance. Attraction was found to be significantly influenced by attitude similarity ($F = 38.03$, $df = 1/76$, $p < .001$), but neither the arousal condition nor the hypothesized interaction yielded a significant F ratio.

It was then decided to divide the experimental group into those most aroused and least aroused by the nonpredictable movie. The dividing point was the median arousal score of 12.5. The mean attraction scores of the high- and low-arousal groups responding to similar and dissimilar strangers are shown in Table 2. A 2 × 2 factorial analysis of variance (summarized in Table 3) indicates not only a significant effect for similarity but also a significant interaction between arousal level and similarity. The direction of the interaction is, however, *opposite* from that which was hypothesized.

Discussion

Not only was the original hypothesis concerning the effect of a nonpredictable movie on attraction to similar and dissimilar strangers not supported, but a significant effect in the reverse direction was found. Given the post hoc nature of the decision to divide subjects into high- and low-arousal groups, it seemed quite possible that the obtained F for the interaction represented a chance finding.

TABLE 3

ANALYSIS OF VARIANCE OF ATTRACTION SCORES OF HIGH- AND LOW-AROUSAL SUBJECTS TOWARD SIMILAR AND DISSIMILAR STRANGERS

Source	*df*	*MS*	*F*
Attitude similarity (A)	1	140.62	28.07***
Effectance arousal (B)	1	4.23	.84
A × B	1	23.44	4.68*
Within	36	5.01	

* $p < .05$.
*** $p < .001$.

Another possibility which was considered was based on speculation about the nature of the arousal state supposedly induced in the subjects. If the nonpredictable movie led not to some generalized effectance arousal but rather to specific concern about the meaning of the film, the relatively smaller differentiation in response to similar and dissimilar strangers may have been a function of the irrelevancy of the attitudinal issues to their current state. That is, those subjects who were confused and uneasy and curious about the movie content were at that moment simply less attentive to or less involved with the content of the attitude scale than subjects who were relatively unaffected by the movie. If this reasoning is correct, the original hypothesis might prove to be accurate providing the stranger's similar and dissimilar attitudes dealt with the movie rather than with topics unrelated to the subject's current arousal state.

III. THE INFLUENCE OF A NONPREDICTABLE MOVIE ON RESPONSE TO SIMILAR AND DISSIMILAR ATTITUDES CONCERNING THE MOVIE

On the basis of the foregoing reasoning, it was hypothesized that, compared to subjects in a low-arousal condition, subjects in an effectance-arousal condition respond more positively to a stranger with similar attitudes and more negatively to a stranger with dissimilar attitudes when the attitude topics are relevant to the arousal stimulus.

Method

The subjects were 162 male and female introductory psychology students at the University of Texas.

In order to provide attitudinal material relevant to the film itself, a 10-item attitude scale was built following the same format as that used in the second investigation. The specific items were based on responses of earlier subjects and dealt with such topics as whether the film was entertaining, whether the central theme was the passage of time, whether the message was an optimistic or a pessimistic one, whether man's primitive or civilized nature was stressed, etc.

The procedures generally followed those of the previous experiment except that it was necessary to give the attitude measure immediately after the movie and then present the subject with a stranger's scale immediately after that. This "standard stranger" technique (Byrne & Clore, 1966) involves the use of a stranger with a random pattern of responses for half of the subjects and a reverse pattern for the other half. The number of similar attitudes for each subject could range from 0 to 10. The procedure then, was movie instructions, movie, attitudes about movie, stranger's attitude scale, judging of stranger, and reaction to movie.

In addition, a new movie (Nonpredictable Movie II) was created by the junior author. This 8-millimeter sound color film includes scenes of a wrecked building, faces on billboards, people crossing a busy street, a silent encounter between a boy and a girl in an open field, chairs marching around a maypole, a toy globe floating in a stream, and other unrelated episodes. The sound track includes a Ravel string quartet, a musical ballad, a woodwind quintet by Poulenc, and a harpsicord etude. Approximately half of the subjects were given Nonpredictable Movie I and half Nonpredictable Movie II. Since neither the effectance arousal scores ($M = 13.17$ and 12.68) nor the attraction results differed for the two movies, the data are combined in the following analysis.

Results

As in the previous study, the sample was dichotomized into high- and low-arousal groups (with a score of 14 as the cutting point). Since the similarity variable in this instance is also a continuum, the subjects were divided into high (.50 to 1.00) and low (.00 to .40) similarity groups. The means and standard deviations of the attraction scores are presented in Table 4. A 2 × 2 factorial analysis of variance (summarized in Table 5) again indicated a significant effect for attitude similarity and a significant interaction between arousal level and similarity. Instead of the hypothesized direction, however, the interaction once again was such that under high-arousal conditions the subjects responded *less* negatively to dissimilar strangers and *less* positively to similar strangers than under low-arousal conditions.

Discussion

It appears that the unexpected results of the second investigation are not attributable to chance or to the content of the attitude items. Before attempting to explain the lack of support for the hypothesis concerning the effect of arousal on response to attitude similarity-dissimilarity, we will examine one more experiment in which a quite different arousal procedure was utilized.

TABLE 4

MEANS AND STANDARD DEVIATIONS OF ATTRACTION SCORES OF HIGH- AND LOW-AROUSAL SUBJECTS TOWARD STRANGERS WITH SIMILAR AND DISSIMILAR ATTITUDES CONCERNING THE MOVIE

Effectance-Arousal level	Proportion of similar attitudes					
	Low[a]		High[b]		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
High	8.76	2.13	9.06	3.09	9.00	2.78
Low	7.46	2.37	9.92	2.38	8.83	2.64
Total	7.80	2.38	9.60	2.63		

[a] .00 to .40.
[b] .50 to 1.00.

IV. THE INFLUENCE OF CONFIRMATION VERSUS DISCONFIRMATION OF A PREDICTION ON RESPONSE TO SIMILAR AND DISSIMILAR ATTITUDES

At the same time that the movie investigations were being conducted, a different experimental procedure was also employed in an effort to manipulate the effectance motive. The general rationale was the same except that instead of confronting subjects with a confusing stimulus, an approach was used in which subjects were asked to predict the outcome of an event in which they were to participate. These predictions were then either

TABLE 5

ANALYSIS OF VARIANCE OF ATTRACTION SCORES OF HIGH- AND LOW-AROUSAL SUBJECTS TOWARD STRANGERS WITH SIMILAR AND DISSIMILAR ATTITUDES CONCERNING THE MOVIE

Source	*df*	*MS*	*F*
Attitude similarity (A)	1	127.02	20.23***
Effectance arousal (B)	1	.24	.04
A × B	1	37.80	6.02*
Within	158	6.28	

Note.—In all analyses which involved disproportional cell frequencies the correction described by Wert, Neidt, and Ahmann (1954) was employed.
* $p < .05$.
*** $p < .001$.

confirmed or disconfirmed. With disconfirmation assumed to arouse the effectance motive, the original hypothesis concerning response to similar and dissimilar strangers following arousal was identical to that proposed in the movie experiments.

Method

The subjects consisted of over 100 male and female students enrolled in introductory psychology and sociology classes at the University of Texas. Early in the semester, they were given the same 12-item attitude survey utilized in the first nonpredictable movie investigation.

At a later experimental session, an experimenter informed each group that they were being asked to participate in a university-wide testing program before serving as subjects in the previously scheduled experiment. They were told:

> The Michigan Intelligence Scale is a test of intelligence and academic ability that has been given to over 4,000 undergraduates at a number of colleges and universities in various parts of the country. We are now in the process of giving the test here to see how students at the University of Texas compare with undergraduates at other institutions. Our Texas sample will be drawn from a number of different classes and the test scores placed on IBM cards along with other data such as grade point average, entrance examination scores, etc.
>
> When I pass out the test, please turn it face down and do not begin until I give the signal. Make all responses on the IBM answer sheet; do not mark on the test booklets. Incidentally, it is better to guess than to leave an item blank. You may use scratch paper for the mathematics section if you wish. You have 20 minutes in which to complete the test. Begin.

After the 20-minute test period, they were told:

> Stop.
>
> Even though our assistants will recheck the tests later, I thought that you might be interested in getting some immediate feedback. I will read the correct answers so that you can score your own test. Place a check mark beside each correct answer. Write the total number of correct answers by the number 1 in the space marked "scores." I will write the percentiles on the board so that you will be able to compare your performance with that of undergraduates at other universities. As you probably know, percentile scores simply tell you where you stand in relationship to others who have already taken the test. For example, the 80th percentile means that 20% of those who have taken the test did better than you, and 80% did not do as well as you. The 40th percentile means that 60% of those who have taken the test did better than you and 40% did not do as well as you. Before I read the answer key, please write down your best guess as to your percentile score in relation to other undergraduates by the number 2 on the answer sheet.

The intelligence test consisted of 25 multiple-choice items involving analogies, vocabulary, information, and mathematics. Actually, two forms of the test were constructed—one relatively easy and one relatively difficult—and both were used at each experimental session. They were arranged in such a way that when the answer key was read aloud, it was applicable to either form. An example of parallel items follows:

Form D	Form E
CATAMARAN is to RAFT as TERMAGANT is to	FISH is to FIN as BIRD is to
a. shrew	a. wing
b. grisette	b. feet
c. benedit	c. beak
d. spinster	d. feathers

Initially it was hoped that a sufficient number of subjects taking each form would make high and low estimates of their ability so that predictions could be disconfirmed in either direction. Actually, almost all subjects predict that they will fall at or above the fiftieth percentile, so that confirmation was primarily confined to those receiving the easy form and disconfirmation to those receiving the difficult form. The bogus percentile information was based on a pilot study in which subjects took both forms of the test.

After the feedback concerning percentiles, a second experimenter took over the group for an "interpersonal judgment" study. The instructions, attitude scales, and the measure of attraction have been described previously. The proportion of similar attitudes expressed by the stranger ranged from .00 to 1.00. Immediately after the experiment, the subjects were told in some detail about the nature and purpose of the "intelligence test."

Results

The bogus percentile rankings which subjects obtained on the difficult form ranged from 5 to 35, and on the easy form from 64 to 94. Because a few of the low-scoring subjects expected not to do very well and because a few of the high-scoring subjects expected to perform somewhat better than the percentile information indicated, subjects were eliminated who did not clearly conform to the desired experimental manipulation. A few other subjects were discarded because of mistakes in their self-scoring or in translating their score into a percentile. The final sample consisted of 70 subjects with 35 in each arousal condition.

The means and standard deviations of the attraction scores are presented in Table 6. The 2 × 2 factorial analysis of variance which is summarized in Table 7 indicated that attraction was significantly influenced by attitude similarity and that there was a significant interaction effect. Just as in the two investigations with the nonpredictable movies, the experimental conditions designed to arouse the effectance motive had the opposite effect on attraction from that which was predicted.

Discussion

The experimental findings indicate a highly consistent effect of the arousal conditions (two different nonpredictable movies and disconfirmation of prediction about performance) on the relationship between attitude similarity (on topics either related or not related to the immediate situation) and attraction. While the empirical findings appear to be on firm ground, the theoretical interpretation of them is another question.

First, it should be noted that on purely logical grounds a convincing case could be made for giving the experimental manipulations a quite different label other than effectance arousal—anxiety arousal, ambiguity, failure, ego threat, etc. The evaluation of the adequacy of the label applied to any operation must ultimately rest on the nature of the empirical relationships with other stimulus and response variables. Given the failure of the original hypotheses about the consequents of effectance arousal, it would be perfectly appropriate to explore some of these other possibilities. One difficulty should be noted by anyone who makes such an attempt, however. The fact that the arousal conditions interact with the similarity-dissimilarity conditions means that any explanation must account for some of the aroused subjects responding *more positively* and some responding *more negatively* to strangers than do the nonaroused subjects. This would suggest that something more must be involved than simply unpleasant affect. Perhaps a reexamination of effectance motivation would prove more fruitful than a search for an alternate explanation.

TABLE 6

MEANS AND STANDARD DEVIATIONS OF ATTRACTION SCORES OF HIGH– AND LOW–AROUSAL SUBJECTS IN INTELLIGENCE TEST SITUATION TOWARD STRANGERS WITH SIMILAR AND DISSIMILAR ATTITUDES

Effectance-Arousal level	Proportion of similar attitudes					
	Low[a]		High[b]		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
High	7.71	1.48	9.57	2.63	8.83	2.42
Low	6.43	1.84	10.71	2.23	9.00	2.96
Total	7.07	1.79	10.14	2.50		

[a] .00 to .33.
[b] .50 to 1.00.

If the nonpredictable movies and the disconfirmation feedback are tentatively assumed to be effectance arousing, it is quite clear that arousal does not lead to an increase in the similarity-attraction relationship. Rather, there is a dampening effect. It is possible that such experimental manipulations are raising drive level to such a point that it is disruptive. In the initial discussion of effectance it was suggested that a highly aroused state of this motive should characterize very young children and psychotics. Further, the operations utilized as an index of the motive include self-ratings of confusion, unreality, and a dream-like state. Thus, a high level of arousal may simply disorient the subject and leave him less sure of his attitudes, less discriminating of the attitudes of others, and perhaps less inclined to express strongly positive or strongly negative evaluations of others. Among the many research possibilities suggested by this formulation is the comparison of schizophrenics and normals in their response to attitude similarity-dissimilarity. One of the characteristics of schizophrenia has been described as dedifferentiation;

TABLE 7

ANALYSIS OF VARIANCE OF ATTRACTION SCORES OF HIGH– AND LOW–AROUSAL SUBJECTS IN INTELLIGENCE–TEST SITUATION TOWARD STRANGERS WITH SIMILAR AND DISSIMILAR ATTITUDES

Source	*df*	*MS*	*F*
Attitude similarity (A)	1	158.49	31.89***
Effectance arousal (B)	1	.52	.10
A × B	1	24.76	4.98*
Within	66	4.97	

* $p < .05$.
*** $p < .001$.

> The distinction between self and nonself (other persons and things), the differentiation between a world of independent objects and one's attitudes and expectations concerning such objects, the distinction between the meanings one feels and seeks to convey and the semantic values of conventional words, the polarity between symbolic vehicles and objects—distinctions all gradually attained during the course of ontogenesis—appear to break down, to varying degrees, in schizophrenic states [Werner & Kaplan, 1963, p. 254].

In terms of the effectance motive, schizophrenics should be characterized by an arousal state higher than that of normals. If the experimental arousal is analogous to this, one would expect schizophrenics to respond like our high-arousal groups in terms of a dampened similarity-attraction relationship. In an investigation just completed by William Griffitt and William Hudgins at the Austin State Hospital,[4] this is precisely what was found. The same linear relationship which is found for college students (Byrne & Clore, 1966) is found to hold in quite different samples including elementary school children (Byrne & Griffitt, 1966) and members of the Job Corps. For schizophrenics, a linear relationship between similarity and attraction is also found, but the slope of the line is significantly less than in these other groups ($F = 11.93$, $df = 3/39$, $p < .001$). That is, these psychotic patients respond like the highly aroused subjects in our experimental groups.

At the moment, then, we have progressed from a theoretical interpretation of the attitude-attraction findings in terms of effectance as an intervening variable to a set of operations designed to arouse this motive, a set of operations designed to measure motive arousal, and some unexpected effects of high arousal.

At this point, it might prove informative to return to the question of attitude items and effectance arousal.

V. ATTITUDE SIMILARITY-DISSIMILARITY AND EFFECTANCE AROUSAL

It has been proposed that similar and dissimilar attitude statements affect attraction responses and also responses in a learning situation because such attitudinal material arouses the effectance motive. Further, in the preceding discussion it was proposed that the arousal effects of the nonpredictable movies and the intelligence test disconfirmation were so great that there was a disruption of the tendency to discriminate between and respond differentially to agreeing and disagreeing strangers. It would be hypothesized that similar and dissimilar attitude statements arouse the effectance motive more than do neutral stimuli and less than do nonpredictable stimuli.

TABLE 8

MEANS AND STANDARD DEVIATIONS OF EFFECTANCE AROUSAL SCORES IN RESPONSE TO CONTROL MOVIE, ATTITUDE ITEMS, AND NONPREDICTABLE MOVIE

	Neutral control movie	Attitude scale items	Nonpredictable movie
M	7.82	9.52	12.73
SD	1.97	2.15	3.66

Method

A total of 46 students in sections of English and history courses at the University of Texas were given an attitude scale; approximately half took a 56-item scale and half a 12-item scale. Afterward, these subjects took part in an "interpersonal judgment experiment," in which they were to judge a stranger on the basis of his or her attitudes on a series of issues. The proportion of the stranger's attitudes which were similar to those of the subject varied from .25 to 1.00.

An additional step in the experimental procedure was the administration of the Movie Reaction Scale described earlier. The title and instructions were altered in order to ask the subjects how they felt "while reading the stranger's attitudes." For comparison purposes, 95 additional students were shown either the *Life in Morocco* control movie or the first nonpredictable movie and then were given the Movie Reaction Scale.

Results

In Table 8 are shown the means and standard deviations of the effectance-arousal scores of subjects in the three experimental conditions. The mean arousal scores of the control and nonpredictable movie groups are quite close to those reported in the first experiment. It may be seen that, as hypothesized, the attitude statements are more arousing than the neutral movie and less arousing than the nonpredictable movie. A simple one-way analysis of variance indicated that these differences were statistically significant ($F = 37.17$, $df = 2/138$, $p < .001$). With respect to the specific hypotheses, arousal was found to be significantly greater in the attitude group than in the control group ($t = 3.78$, $df = 84$, $p < .001$), and significantly less in the attitude group than in the nonpredictable movie group ($t = 5.18$, $df = 99$, $p < .001$).

---

[4] The authors wish to express their appreciation to Superintendent C. R. Miller and Chief Psychologist Robert Reinehr of the Austin State Hospital for their cooperation and assistance in conducting this portion of the investigation.

Two additional questions were explored. If attitude items do, in fact, arouse the effectance motive, it would seem that the number of attitude items would be positively related to amount of arousal. Comparison of the mean arousal scores of the group receiving the 12-item scale ($M = 9.62$) with that of the group receiving the 56-item scale ($M = 9.47$) indicated a nonsignificant difference.

It also seemed probable that the proportion of items which were similar to and dissimilar from the subject's views should be related to degree of arousal. That is, exposure to attitudinal material and to the possibility of disagreement should in itself be arousing; as the subject discovers items on which the stranger holds a position dissimilar from his own, effectance arousal should increase as a positive function of the number of such items relative to the number of items on which there is agreement. Dividing the 46 subjects who received attitude scales into the 23 with relatively similar strangers (.58 to 1.00 similarity) and the 23 with relatively dissimilar strangers (.25 to .50), effectance-arousal scores were compared. The mean arousal score of the former group was found to be 8.83 and of the latter group, 10.22; the difference between these means is a statistically significant one ($t = 2.28$, $df = 44$, $p < .05$). Not only does attitudinal material arouse the effectance motive, but arousal is differentially influenced by similar and dissimilar attitude statements.

Discussion

It was proposed in investigations II, III, and IV that effectance arousal would increase or heighten the similarity-attraction relationship. The opposite effect was found; for those individuals who indicated a high arousal state, there was a decrease in the similarity-attraction relationship. It has been proposed that this effect was the result of a disorganizing or dedifferentiating or disorienting effect of extreme arousal.

An alternate approach to the original hypotheses would be to measure arousal in the attitude situation by means of the reaction scale and to predict differential effects on attraction for relatively aroused and nonaroused subjects. The difference between this approach and that reported earlier, of course, is that the arousal levels are lower than those created by the nonpredictable movies, and presumably lower than those created by the intelligence-test manipulation.

The subjects in Investigation V who received the attitude scale of a stranger were supplemented by an additional group of subjects who went through the same procedure of judging a stranger on the basis of his attitudes and of indicating their arousal on the reaction scale. Then the 42 subjects who received a relatively dissimilar stranger (.15 to .34 similarity) and the 49 subjects who received a relatively similar stranger (.85 to 1.00 similarity) were [illegible] divided into high- and low-arousal groups on the basis of their effectance-arousal scores (5–9 for low group, 10–14 for high group). The mean attraction scores of these groups are shown in Table 9. It may be seen that the direction of the interaction is that which was originally hypothesized. That is, as arousal increases, there is a more negative response to a dissimilar stranger and a more positive response

TABLE 9

MEANS AND STANDARD DEVIATIONS OF ATTRACTION SCORES OF HIGH- AND LOW-AROUSAL SUBJECTS IN INTERPERSONAL JUDGMENT SITUATION TOWARD STRANGERS WITH SIMILAR AND DISSIMILAR ATTITUDES

Effectance-Arousal level	Proportion of similar attitudes					
	Low[a]		High[b]		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
High (10–14)	6.00	2.50	10.96	1.97	8.53	3.34
Low (5–9)	7.50	2.13	10.31	2.64	9.09	2.80
Total	6.71	2.45	10.61	2.37		

[a] .15 to .34.
[b] .85 to 1.00.

TABLE 10

MEANS AND STANDARD DEVIATIONS OF ATTRACTION SCORES OF MODERATELY AROUSED AND NONAROUSED SUBJECTS TOWARD SIMILAR AND DISSIMILAR STRANGERS

Effectance-Arousal level	Proportion of similar attitudes					
	Low[a]		High[b]		Total	
	M	*SD*	*M*	*SD*	*M*	*SD*
Moderate (9–12)	6.29	2.08	11.73	.91	9.36	3.10
Nonaroused (5–8)	7.36	2.83	10.49	1.70	9.23	2.70
Total	6.93	2.60	10.95	1.58		

[a] .10 to .20.
[b] .70 to .90.

to a similar stranger. Analysis of variance indicates, however, that this interaction does not quite reach the .05 level of significance ($F = 3.83$, $df = 1/87$). The predicted difference in attraction between high- and low-arousal groups toward a dissimilar stranger is a significant one ($t = 2.03$, $df = 40$, $p < .05$). Thus, support for the hypothesized relationship is somewhat tenuous.

Such findings led us to seek further an explanation for the failure of the experimental movies to influence response to similar and dissimilar strangers in the expected way. The dampening effect of high arousal was consistently observed, but what about the group we have labeled "low arousal?" A reexamination of the effectance scores of the subjects in investigations II and III was made. The subjects who were shown the neutral control movie had a mean arousal score of 7.20. The "low-arousal" subjects in the two movie investigations yielded a mean arousal score of 10.71, while the mean of the high-arousal subjects was 16.06. Thus, even the lower half of the distribution of those subjects who viewed the nonpredictable movies, was *more aroused* than the subjects in the control group. There are actually three levels of arousal represented in those investigations—low (control group), moderate (lower half of experimental groups), and high (upper half of experimental groups). The influence of moderate levels of effectance arousal could be ascertained by comparing the control subjects with the moderately aroused experimental subjects.

TABLE 11

ANALYSIS OF VARIANCE OF ATTRACTION SCORES OF MODERATELY AROUSED AND NONAROUSED SUBJECTS TOWARD SIMILAR AND DISSIMILAR STRANGERS

Source	*df*	*MS*	*F*
Attitude similarity (A)	1	397.85	96.80***
Effectance arousal (B)	1	1.68	.41
A × B	1	31.06	7.56**
Within	97	4.11	

** $p < .01$.
*** $p < .001$.

In order to have a sufficiently large *N*, comparable subjects in the two movie studies were combined to form four groups with a total of 101 individuals. The control subjects were those who had viewed *Life in Morocco*. The moderate-arousal subjects had viewed either the first or second nonpredictable movie; their effectance-arousal scores ranged from 9 to 12. The attitude-similarity conditions were limited to relatively low (.10 to .20 similarity) and relatively high (.70 to .90 similarity) agreement. The means and standard deviations of the attraction scores are shown in Table 10. The direction of the interaction is that predicted originally, and as may be seen in Table 11, is significant at the .01 level.

It would appear that the initial failure to find a difference between the attraction responses of the experimental and control subjects was brought about by the quite different effects of moderate and high drive arousal. Compared to a nonarousal control situation,

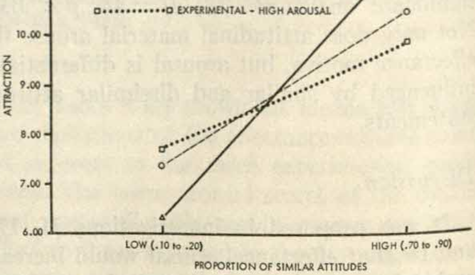


FIG. 1. Attraction toward a stranger as an interactive function of proportion of similar attitudes and level of effectance arousal.

moderate effectance arousal leads to an increased negative response toward a dissimilar stranger and an increased positive response toward a similar one; high affectance arousal influences attraction in just the reverse manner. These relationships are seen most clearly in Figure 1.

In manipulating the motivational state which is presumed to mediate the attitude-attraction relationship, the level of arousal apparently is a crucial variable.

CONCLUSIONS

In the foregoing series of experiments, an attempt was made to investigate some of the implications of a motivational interpretation of the effects of similar and dissimilar attitudes on attraction. By equating the hypothesized motive state with similar motivational constructs proposed by a variety of investigators in widely ranging contexts, it was possible to suggest operations, independent of the attitude stimuli and attraction responses, designed to arouse and to indicate the degree of arousal of the effectance motive.

Several findings seem relatively clear. First, the nonpredictable movies arouse the motive state as measured by the reaction scale. This arousal is not dependent on the specific content of a specific movie as shown by the nearly identical results obtained with movies I and II. Further, it appears that the arousal effects of these stimulus conditions are a function of the ambiguous and confusing elements rather than a more general effect such as emotional arousal as shown by the comparison of the cataract operation movie with the nonpredictable movies.

Second, attitude statements arouse the effectance motive at a level intermediate between a neutral stimulus and the nonpredictable movies. In addition, the amount of arousal increases as the proportion of dissimilar attitudes increases. If one can tentatively accept the motivation measure as a valid one, this finding constitutes a direct confirmation of the hypothetical intervening construct in the attitude-attraction relationship. Furthermore, at this level of arousal (unlike the higher levels evoked by the movies) there is some evidence that the magnitude of arousal is negatively related to attraction toward a dissimilar stranger and positively related to attraction toward a similar stranger.

Third, a high level of effectance arousal as brought about by both movies and by a quite different manipulation in the intelligence-test situation has a significant effect on response to similar and dissimilar strangers. Contrary to the original hypothesis, high arousal resulted in a less negative response to dissimilarity and a less positive response to similarity. This unanticipated finding led to the proposal that high levels of effectance arousal are disorganizing or disorienting and hence result in decreased sensitivity to environmental variation. The finding that schizophrenic patients respond to attitude similarity-dissimilarity in the same way that our highly aroused subjects do is at least compatible with the general proposition.

A reexamination of the movie data indicates that, compared to a neutral situation, a moderate level of effectance arousal results in a more positive response toward a similar stranger and a more negative response toward a dissimilar stranger. Thus, as effectance level increases, the attitude-attraction relationship is initially increased, but, as higher levels are reached, the attitude-attraction relationship is less than in a neutral situation.

A final cautionary note should be added. The theoretical speculations which led to the present series of studies consisted of a tenuous amalgam of observations, anecdotes, and empirical data. The correspondence between the experimental situations designed to arouse effectance motivation or the response measure designed to provide an index of such arousal and the corresponding theoretical constructs necessarily depends on one's subjective evaluation. The obtained findings, however, can be made to fit reasonably well within the proposed framework. Obviously, other labels may legitimately be appended to this same set of operations and different theoretical explanations may be offered to account for the obtained relationships. The reciprocal relationship between data and theory and the transitory nature of the latter are sufficiently familiar that one's effectance motive need not become aroused. For the present, at any rate, the work reported here does provide empirical support for a hypothetical motivational construct which mediates the relationship between attitude statements and attraction, and in addition it suggests the generality of these phenomena considerably beyond this specific experimental situation.

SUMMARY

The empirically established relationship between attitude similarity-dissimilarity and attraction has been interpreted by both cognitive and reinforcement theorists in motivational terms. In the present paper, the motive aroused by attitudinal dissimilarity was tentatively equated with the type of motive labeled in other contexts as the need for cognition, need to know, desire for certainty, etc. White's term, *effectance motive,* was used to encompass all of these constructs. A series of experiments was carried out to explore the proposal that the effectance motive is aroused by attitude material and that differential drive level influences attraction responses.

As an independent arousal stimulus, a nonpredictable movie was constructed, and a reaction scale was built as an index of effectance arousal. In the first investigation ($N = 120$) the nonpredictable movie was found to yield higher effectance arousal scores than either a neutral movie or a predictable but emotion-arousing movie depicting a cataract operation. In the second investigation ($N = 80$), subjects saw either the nonpredictable movie or a control movie and afterwards responded to a stranger who was either relatively similar to or dissimilar from themselves in attitudes. Attraction was found to vary as a function of attitude similarity but not as a function of arousal condition. Then, the experimental group was divided into high- and low-arousal subjects. A significant interaction was found in a direction opposite to that hypothesized; the high-arousal group was less negative toward a dissimilar stranger and less positive toward a similar stranger than was the low-arousal group. Investigation III ($N = 162$) was like the second investigation except that two different nonpredictable movies were employed and the attitude items dealt with the movie itself rather than with other topics. Again, a significant interaction was found with high-arousal subjects showing a less marked attitude-attraction relationship than low-arousal subjects. In Investigation IV ($N = 70$) a different arousal procedure was used. Individuals predicted their performance on an intelligence test and then were given feedback which either confirmed or disconfirmed their prediction; the latter information was assumed to constitute arousal. In responding to a stranger afterward, precisely the same significant interaction was found as in the two movie experiments. In a fifth investigation ($N = 186$), attitude items were found to arouse the effectance motive to a greater extent than a neutral stimulus and less than the nonpredictable movies. Further, the degree of arousal was a positive function of the proportion of dissimilar attitudes. In addition, as arousal increases, response to a dissimilar stranger becomes more negative. On the basis of these findings, the movie data were reexamined. It was found that moderate drive arousal does heighten the attitude-attraction relationship compared to a nonaroused situation while high drive arousal has the opposite effect.

It was concluded that the nonpredictable movies arouse the effectance motive as measured by the rating scale, attitude statements also arouse the effectance motive but less than do the movies, and a moderate level of effectance arousal increases the slope of the attitude-attraction function while a high arousal level decreases the slope of the function.

REFERENCES

BERLYNE, D. E. The arousal and satiation of perceptual curiosity in the rat. *Journal of Comparative and Physiological Psychology,* 1955, **48,** 238–246.

BRIM, O. G., JR., & HOFF, D. B. Individual and situational differences in desire for certainty. *Journal of Abnormal and Social Psychology,* 1957, **54,** 225–229.

BYRNE, D. Interpersonal attraction and attitude similarity. *Journal of Abnormal and Social Psychology,* 1961, **62,** 713–715.

BYRNE, D. Response to attitude similarity-dissimilarity as a function of affiliation need. *Journal of Personality,* 1962, **30,** 164–177.

Byrne, D. *An introduction to personality: A research approach*. Englewood Cliffs, N. J.: Prentice-Hall, 1966.

Byrne, D., & Clore, G. L., Jr. Predicting interpersonal attraction toward strangers presented in three different stimulus modes. *Psychonomic Science*, 1966, **4**, 239–240.

Byrne, D., & Griffitt, W. A developmental investigation of the law of attraction. *Journal of Personality and Social Psychology*, 1966, **4**, 699–702.

Byrne, D., & Nelson, D. Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology*, 1965, **1**, 659–663.

Byrne, D., Nelson, D., & Reeves, K. Effects of consensual validation and invalidation on attraction as a function of verifiability. *Journal of Experimental Social Psychology*, 1966, **2**, 98–107.

Byrne, D., & Rhamey, R. Magnitude of positive and negative reinforcements as a determinant of attraction. *Journal of Personality and Social Psychology*, 1965, **2**, 884–889.

Byrne, D., Young, R. K., & Griffitt, W. The reinforcement properties of attitude statements. *Journal of Experimental Research in Personality*, 1966, **1**, 266–276.

Carlsmith, J. M., & Aronson, E. Some hedonic consequences of the confirmation and disconfirmation of expectancies. *Journal of Abnormal and Social Psychology*, 1963, **66**, 151–156.

Clore, G. L., Jr. Discrimination learning as a function of awareness and magnitude of attitudinal reinforcement. Unpublished doctoral dissertation, University of Texas, 1966.

Cohen, A., Stotland, E., & Wolfe, D. M. An experimental investigation of need for cognition. *Journal of Abnormal and Social Psychology*, 1955, **51**, 291–297.

D'Amato, M. R., & Gumenik, W. E. Some effects of immediate versus randomly delayed shock on an instrumental response and cognitive processes. *Journal of Abnormal and Social Psychology*, 1960, **60**, 64–67.

Dember, W. N., & Earl, R. W. Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review*, 1957, **64**, 91–96.

Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

Elliott, R. Effects of uncertainty about the nature and advent of a noxious stimulus (shock) upon heart rate. *Journal of Personality and Social Psychology*, 1966, **3**, 353–356.

Festinger, L. Informal social communication. *Psychological Review*, 1950, **57**, 271–282.

Festinger, L. A theory of social comparison processes. *Human Relations*, 1954, **7**, 117–140.

Gerard, H. Emotional uncertainty and social comparison. *Journal of Abnormal and Social Psychology*, 1963, **66**, 568–573.

Gergen, K. J., & Jones, E. E. Mental illness, predictability, and affective consequences as stimulus factors in person perception. *Journal of Abnormal and Social Psychology*, 1963, **67**, 95–104.

Golightly, C., & Byrne, D. Attitude statements as positive and negative reinforcements. *Science*, 1964, **146**, 798–799.

Hebb, D. O. *Organization of behavior: A neuropsychological theory*. New York: Wiley, 1949.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Jones, A., Wilkinson, H. J., & Braden, I. Information deprivation as a motivational variable. *Journal of Experimental Psychology*, 1961, **62**, 126–137.

Kelly, G. A. *The psychology of personal constructs*. New York: Norton, 1955.

Lanzetta, J. T., & Driscoll, J. M. Preference for information about an uncertain but unavoidable outcome. *Journal of Personality and Social Psychology*, 1966, **3**, 96–102.

Leuba, C. Toward some integration of learning theories: The concept of optimal stimulation. *Psychological Reports*, 1955, **1**, 27–33.

Maddi, S. Affective tone during environmental regularity and change. *Journal of Abnormal and Social Psychology*, 1961, **62**, 338–345.

McReynolds, P. A restricted conceptualization of human anxiety and motivation. *Psychological Reports*, 1956, **2**(Monogr. Suppl. No. 6), 293–312.

Munsinger, H., & Kessen, W. Uncertainty, structure, and preference. *Psychological Monographs*, 1964, **78**(9, Whole No. 586).

Nelson, D. The effect of differential magnitude of reinforcement on interpersonal attraction. Unpublished doctoral dissertation, University of Texas, 1965.

Newcomb, T. M. An approach to the study of communicative acts. *Psychological Review*, 1953, **60**, 393–404.

Newcomb, T. M. Individual systems of orientation. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 3. *Formulations of the person and the social context*. New York: McGraw-Hill, 1959. Pp. 384–422.

Newcomb, T. M. *The acquaintance process*. New York: Holt, Rinehart & Winston, 1961.

Pervin, L. A. The need to predict and control under conditions of threat. *Journal of Personality*, 1963, **31**, 570–587.

Schwartz, M. S. Prediction of individual differences in the arousal of the effectance motive and in interpersonal attraction toward a stranger identified with the arousal of the effectance motive. Unpublished doctoral dissertation, University of Texas, 1966.

Welker, W. L. Some determinants of play and exploration in chimpanzees. *Journal of Comparative and Physiological Psychology*, 1956, **49**, 84–89.

Werner, H., & Kaplan, B. *Symbol formation*. New York: Wiley, 1963.

Wert, J. E., Neidt, C. O., & Ahmann, J. S. *Statistical methods in educational and psychological research*. New York: Appleton-Century-Crofts, 1954.

WHITE, R. W. Motivation reconsidered: the concept of competence. *Psychological Review,* 1959, **66,** 297–333.

WOODWORTH, R. S. *Dynamics of behavior.* New York: Holt, 1958.

ZIMBARDO, P. G., & MILLER, N. E. Facilitation of exploration by hunger in rats. *Journal of Comparative and Physiological Psychology,* 1958, **51,** 43–46.

APPENDIX A

REACTION SCALE

Name________________________

How did you feel while you were watching the movie? (reading the stranger's attitudes?)

1. Entertained (check one)

____Not at all entertained
____Slightly entertained
____Moderately entertained
____Entertained
____Quite entertained

2. Disgusted (check one)

____Not at all disgusted
____Slightly disgusted
____Moderately disgusted
____Disgusted
____Extremely disgusted

3. Unreality (check one)

____Strong feelings of unreality
____Feelings of unreality
____Moderate feelings of unreality
____Slight feelings of unreality
____No feelings of unreality at all

4. Anxious (check one)

____Not at all anxious
____Slightly anxious
____Moderately anxious
____Anxious
____Extremely anxious

5. Dreaming (check one)

____Very similar to feelings I have when I'm dreaming
____Similar to feelings I have when I'm dreaming
____Moderately similar to feelings I have when I'm dreaming
____Slightly similar to feelings I have when I'm dreaming
____Not at all similar to feelings I have when I'm dreaming

6. Bored (check one)

____Extremely bored
____Bored
____Moderately bored
____Slightly bored
____Not at all bored

7. Uneasy (check one)

____Not at all uneasy
____Slightly uneasy
____Moderately Uneasy
____Uneasy
____Quite uneasy

8. Confused (check one)

____Not at all confused
____Slightly confused
____Moderately confused
____Confused
____Quite confused

9. Others' thoughts (check one)

____Strong desire to know what others thought
____Desire to know what others thought
____Moderate desire to know what others thought
____Slight desire to know what others thought
____No desire to know what others thought at all

(Received August 12, 1966)